



**University of  
Zurich**<sup>UZH</sup>

**Zurich Open Repository and  
Archive**

University of Zurich  
University Library  
Strickhofstrasse 39  
CH-8057 Zurich  
[www.zora.uzh.ch](http://www.zora.uzh.ch)

---

Year: 2017

---

## **Environmental DNA metabarcoding: Transforming how we survey animal and plant communities**

Deiner, Kristy ; Bik, Holly M ; Mächler, Elvira ; Seymour, Mathew ; Lacoursière-Roussel, Anaïs ;  
Altermatt, Florian ; Creer, Simon ; Bista, Iliana ; Lodge, David M ; de Vere, Natasha ; Pfrender,  
Michael E ; Bernatchez, Louis

**Abstract:** The genomic revolution has fundamentally changed how we survey biodiversity on earth. High-throughput sequencing (“HTS”) platforms now enable the rapid sequencing of DNA from diverse kinds of environmental samples (termed “environmental DNA” or “eDNA”). Coupling HTS with our ability to associate sequences from eDNA with a taxonomic name is called “eDNA metabarcoding” and offers a powerful molecular tool capable of noninvasively surveying species richness from many ecosystems. Here, we review the use of eDNA metabarcoding for surveying animal and plant richness, and the challenges in using eDNA approaches to estimate relative abundance. We highlight eDNA applications in freshwater, marine and terrestrial environments, and in this broad context, we distill what is known about the ability of different eDNA sample types to approximate richness in space and across time. We provide guiding questions for study design and discuss the eDNA metabarcoding workflow with a focus on primers and library preparation methods. We additionally discuss important criteria for consideration of bioinformatic filtering of data sets, with recommendations for increasing transparency. Finally, looking to the future, we discuss emerging applications of eDNA metabarcoding in ecology, conservation, invasion biology, biomonitoring, and how eDNA metabarcoding can empower citizen science and biodiversity education.

DOI: <https://doi.org/10.1111/mec.14350>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-144367>

Journal Article

Published Version









The following work is licensed under a Creative Commons: Attribution 4.0 International (CC BY 4.0) License.

Originally published at:

Deiner, Kristy; Bik, Holly M; Mächler, Elvira; Seymour, Mathew; Lacoursière-Roussel, Anaïs; Altermatt, Florian; Creer, Simon; Bista, Iliana; Lodge, David M; de Vere, Natasha; Pfrender, Michael E; Bernatchez, Louis (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26(21):5872-5895.

DOI: <https://doi.org/10.1111/mec.14350>

# Environmental DNA metabarcoding: Transforming how we survey animal and plant communities

Kristy Deiner<sup>1</sup>  | Holly M. Bik<sup>2</sup>  | Elvira Mächler<sup>3,4</sup> | Mathew Seymour<sup>5</sup> |  
 Anaïs Lacoursière-Roussel<sup>6</sup>  | Florian Altermatt<sup>3,4</sup>  | Simon Creer<sup>5</sup>  | Iliana Bista<sup>5,7</sup> |  
 David M. Lodge<sup>1</sup> | Natasha de Vere<sup>8,9</sup>  | Michael E. Pfrender<sup>10</sup> | Louis Bernatchez<sup>6</sup>

<sup>1</sup>Atkinson Center for a Sustainable Future, Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA

<sup>2</sup>Department of Nematology, University of California, Riverside, CA, USA

<sup>3</sup>Eawag, Swiss Federal Institute of Aquatic Science and Technology, Department of Aquatic Ecology, Dübendorf, Switzerland

<sup>4</sup>Department of Evolutionary Biology and Environmental Studies, University of Zurich, Zürich, Switzerland

<sup>5</sup>Molecular Ecology and Fisheries Genetics Laboratory, School of Biological Sciences, Environment Centre Wales Building, Bangor University, Bangor, Gwynedd, UK

<sup>6</sup>IBIS (Institut de Biologie Intégrative et des Systèmes), Université Laval, Québec, QC, Canada

<sup>7</sup>Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, UK

<sup>8</sup>Conservation and Research Department, National Botanic Garden of Wales, Llanarthne, Carmarthenshire, UK

<sup>9</sup>Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK

<sup>10</sup>Department of Biological Sciences and Environmental Change Initiative, University of Notre Dame, Notre Dame, IN, USA

## Correspondence

Kristy Deiner, Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA.  
 Email: alpinedna@gmail.com

## Funding information

U.S. Department of Defense, Grant/Award Number: W912HQ-12-C-0073 (RC-2240); National Oceanic and Atmospheric Administration; National Science Foundation, Grant/Award Number: EF-1427157, DBI-1262480; European Cooperation in Science and Technology, Grant/Award Number: CA15219; Swiss National Science Foundation, Grant/Award Number: PP00P3\_150698; Natural Environment Research Council, Grant/Award Number: NBAF824 2013-14, NE/N003756/1, NE/N006216/1; Knowledge Economy Skills Scholarship

## Abstract

The genomic revolution has fundamentally changed how we survey biodiversity on earth. High-throughput sequencing ("HTS") platforms now enable the rapid sequencing of DNA from diverse kinds of environmental samples (termed "environmental DNA" or "eDNA"). Coupling HTS with our ability to associate sequences from eDNA with a taxonomic name is called "eDNA metabarcoding" and offers a powerful molecular tool capable of noninvasively surveying species richness from many ecosystems. Here, we review the use of eDNA metabarcoding for surveying animal and plant richness, and the challenges in using eDNA approaches to estimate relative abundance. We highlight eDNA applications in freshwater, marine and terrestrial environments, and in this broad context, we distill what is known about the ability of different eDNA sample types to approximate richness in space and across time. We provide guiding questions for study design and discuss the eDNA metabarcoding workflow with a focus on primers and library preparation methods. We additionally discuss important criteria for consideration of bioinformatic filtering of data sets, with recommendations for increasing transparency. Finally, looking to the future, we discuss emerging applications of eDNA metabarcoding in ecology, conservation, invasion biology, biomonitoring, and how eDNA metabarcoding can empower citizen science and biodiversity education.

**KEYWORDS**

bioinformatic pipeline, biomonitoring, citizen science, conservation, ecology, eDNA, invasive species, macro-organism, species richness

## 1 | INTRODUCTION

Anthropogenic influences are causing unprecedented changes to the rate of biodiversity loss and, consequently, ecosystem function (Cardinale et al., 2012). Accordingly, we need rapid biodiversity survey tools for measuring fluctuations in species richness to inform conservation and management strategies (Kelly et al., 2014). Multispecies detection using DNA derived from environmental samples (termed “environmental DNA” or “eDNA”) using high-throughput sequencing (“HTS”; Box 1) is a fast and efficient method to survey species richness in natural communities (Creer et al., 2016). Bacterial and fungal taxonomic richness (i.e., richness of microorganisms) is routinely surveyed using DNA metabarcoding and is a powerful complement to conventional culture-based methods (e.g., Caporaso et al., 2011; Tedersoo et al., 2014). Over the last decade, it has been recognized that animal and plant communities can be surveyed in a similar fashion (Taberlet, Coissac, Pompanon, Brochmann, & Willerslev, 2012; Valentini, Pompanon, & Taberlet, 2009).

Many literature reviews summarize how environmental DNA (eDNA) can be used to detect biodiversity, but they focus on single-species detections, richness estimates from community DNA (see Box 1 for definition for how this differs and can be confused with eDNA) or general aspects of using eDNA for the detection of biodiversity in a specific field of study (Table S1). To complement these many recent reviews, here we concentrate on four aspects: a summary of eDNA metabarcoding studies on animals and plants to date, knowns and unknowns surrounding the spatial and temporal scale of eDNA information, guidelines and challenges for eDNA study design (with a specific focus on primers and library preparation) and emerging applications of eDNA metabarcoding in the basic and applied sciences.

## 2 | SURVEYING SPECIES RICHNESS AND RELATIVE ABUNDANCE WITH EDNA METABARCODING

Conventional physical, acoustic and visual-based methods for surveying species richness and relative abundance have been the major ways we observe biodiversity, yet they are not without limitations. For instance, despite highly specialized identification by experts, in some taxonomic groups identification errors are common (Bortolus, 2008; Stribling, Pavlik, Holdsworth, & Leppo, 2008). Conventional physical methods can also cause destructive impacts on the environment and to biological communities (Wheeler, Raven, & Wilson, 2004), making them difficult to apply in a conservation context. Furthermore, when a species' behaviour or size makes it difficult to

survey them (e.g., small-bodied or elusive species), conventional methods can require specialized equipment or species-specific observation times, thus making species richness and relative abundance estimates for entire communities intractable (e.g., many amphibians and reptiles, Erb, Willey, Johnson, Hines, & Cook, 2015; Price, Eskew, Cecala, Browne, & Dorcas, 2012). These reasons highlight the continued need to develop improved ways to survey global biodiversity, and the unique ways eDNA metabarcoding can complement conventional methods.

### 2.1 | Species richness: eDNA metabarcoding compared with conventional methods

Environmental DNA metabarcoding can complement (and overcome the limitations of) conventional methods by targeting different species, sampling greater diversity and increasing the resolution of taxonomic identifications (Table 1). For example, Valentini et al. (2016) demonstrated that, for many different aquatic systems, the number of amphibian species detected using eDNA metabarcoding was equal to or greater than the number detected using conventional methods. When terrestrial haematophagous leeches were used as collectors of eDNA (blood of hosts), endangered and elusive vertebrate species were detected using eDNA metabarcoding and served as a valuable complement to camera trap surveys in a remote geographic region (Schnell, Sollmann, et al. 2015). In plants, Kraaijeveld et al. (2015) demonstrated that eDNA metabarcoding of filtered air samples allowed pollen to be identified with greater taxonomic resolution relative to visual methods.

The ways that eDNA can complement and extend conventional surveys are promising, but the spatial and temporal scale of inference is likely to differ between conventional and molecular methods. For example, in a river, Deiner, Fronhofer, Mächler, Walser, and Altermatt (2016) showed on a site-by-site basis that the eDNA metabarcoding method resulted in higher species detection compared to a conventional physical-capture method (i.e., kicknet sampling; Table 1). However, eDNA in this case may have detected greater species richness at a site not because the species themselves are present, but rather because their DNA has been transported from another location upstream, creating an inference challenge in space and time for eDNA species detections. Therefore, research is needed to understand the complex spatiotemporal dynamics of the various eDNA sample types (Figure 1), which at present we know very little about. In addition, all sampling methods have inherent biases caused by their detection probabilities. Detection probabilities often vary by species, habitat and detection method (e.g., the mesh size of a net or a primer's match to a target DNA sequence) and use of bias-corrected species richness estimators will be important to

**BOX 1 COMMUNITY DNA VERSUS ENVIRONMENTAL DNA METABARCODING OF PLANTS AND ANIMALS****Terms****Environmental DNA (eDNA)**

DNA captured from an environmental sample without first isolating any target organisms (Taberlet, Coissac, Hajibabaei, & Rieseberg, 2012). Traces of DNA can be from faeces, mucus, skin cells, organelles, gametes or even extracellular DNA. Environmental DNA can be sampled from modern environments (e.g., seawater, freshwater, soil or air) or ancient environments (e.g., cores from sediment, ice or permafrost, see Thomsen & Willerslev, 2015).

**Community DNA**

DNA is isolated from bulk-extracted mixtures of organisms separated from the environmental sample (e.g., soil or water).

**Macro-organism environmental DNA**

Environmental DNA originating from animals and higher plants.

**Barcoding**

First defined by Hebert et al. (2003), the term refers to taxonomic identification of species based on single specimen sequencing of diagnostic barcoding markers (e.g., COI, *rbcL*).

**Metabarcoding**

Taxonomic identification of multiple species extracted from a mixed sample (community DNA or eDNA) which have been PCR-amplified and sequenced on a high-throughput platform (e.g., Illumina, Ion Torrent).

**High-throughput sequencing (HTS)**

Sequencing techniques that allow for simultaneous analysis of millions of sequences compared to the Sanger sequencing method of processing one sequence at a time.

**Community DNA metabarcoding**

HTS of DNA extracted from specimens or whole organisms collected together, but first separated from the environmental sample (e.g., water or soil).

**Molecular Operational Taxonomic Unit (MOTU)**

Group identified through use of cluster algorithms and a predefined percentage sequence similarity (e.g., 97%; Blaxter et al., 2005).

Since the inception of high-throughput sequencing (HTS, Margulies et al., 2005), the use of metabarcoding as a biodiversity detection tool has drawn immense interest (e.g., Creer et al., 2016; Hajibabaei et al., 2011). However, there has yet to be clarity regarding what source material is used to conduct metabarcoding analyses (e.g., environmental DNA versus community DNA). Without clarity between these two source materials, differences in sampling, as well as differences in laboratory procedures, can impact subsequent bioinformatics pipelines used for data processing, and complicate the interpretation of spatial and temporal biodiversity patterns. Here, we seek to clearly differentiate among the prevailing source materials used and their effect on downstream analysis and interpretation for environmental DNA metabarcoding of animals and plants compared to that of community DNA metabarcoding.

With community DNA metabarcoding of animals and plants, the targeted groups are most often collected in bulk (e.g., soil, malaise trap or net), and individuals are removed from other sample debris and pooled together prior to bulk DNA extraction (Creer et al., 2016). In contrast, macro-organism eDNA is isolated directly from an environmental material (e.g., soil or water) without prior segregation of individual organisms or plant material from the sample and implicitly assumes that the whole organism is not present in the sample. Of course, community DNA samples may contain DNA from parts of tissues, cells and organelles of other organisms (e.g., gut contents, cutaneous intracellular or extracellular DNA). Likewise, macro-organism eDNA samples may inadvertently capture whole microscopic nontarget organisms (e.g., protists, bacteria). Thus, the distinction can at least partly break down in practice.

Another important distinction between community DNA and macro-organism eDNA is that sequences generated from community DNA metabarcoding can be taxonomically verified when the specimens are not destroyed in the extraction process. Here, sequences can then be generated from voucher specimens using Sanger sequencing. As the samples for eDNA metabarcoding lack whole organisms, no such in situ comparisons can be made. Taxonomic affinities can therefore only be established by directly comparing obtained sequences (or through bioinformatically generated operational taxonomic units (MOTUs)), to sequences that are taxonomically annotated such as NCBI's GenBank nucleotide database (Benson et al., 2013), BOLD (Ratnasingham & Hebert, 2007), or to self-generated

(Continues)

**BOX 1** (Continued)

reference databases from Sanger-sequenced DNA (Olds et al., 2016; Sønstebo et al., 2010; Willerslev et al., 2014). Then, to at least partially corroborate the resulting list of taxa, comparisons are made with conventional physical, acoustic or visual-based survey methods conducted at the same time or compared with historical records from surveys for a location (see Table 1).

The difference in source material between community DNA and eDNA therefore has distinct ramifications for interpreting the scale of inference for time and space about the biodiversity detected. From community DNA, it is clear that the individual species were found in that time and place, but for eDNA, the organism that produced the DNA may be upstream from the sampled location (Deiner & Altermatt, 2014), or the DNA may have been transported in the faeces of a more mobile predatory species (e.g., birds depositing fish eDNA, Merkes, McCalla, Jensen, Gaikowski, & Amberg, 2014) or was previously present, but no longer active in the community and detection is from DNA that was shed years to decades before (Yoccoz et al., 2012). The latter means that the scale of inference both in space and in time must be considered carefully when inferring the presence for the species in the community based on eDNA.

account for these biases when conducting statistical comparisons between the outcomes in measured richness (Gotelli & Colwell, 2011; Olds et al., 2016).

Future methodological comparisons could also benefit from a quantitative ecological approach in the design of sampling by matching sample effort and scope of sampling between eDNA and conventional methods. Multimethod species distribution modelling or site occupancy modelling is one example for how this can be achieved and has been demonstrated in cases comparing qPCR for a single species and conventional methods (Hunter et al., 2015; Rees, Bishop, Middleditch, et al. 2014; Schmelzle & Kinziger, 2016; Schmidt, Kery, Ursenbacher, Hyman, & Collins, 2013), but rarely for eDNA metabarcoding (Ficetola et al., 2015). Thus, we expect the robustness of eDNA metabarcoding to reveal species richness estimates for animals and plants will be improved by coupling distribution or occupancy modelling with studies to determine the scale of inference in space and time for an eDNA sample (Figure 1).

## 2.2 | Species relative abundance: eDNA metabarcoding compared with conventional methods

Estimating a species' relative abundance using eDNA metabarcoding is an intriguing possibility. Here, we focus on the evidence from animals in aquatic systems. Controlled studies based on the detection of a single animal species in small ecosystems, such as in aquaria and mesocosms (e.g., Minamoto, Yamanaka, Takahara, Honjo, & Zi, 2012; Moyer, Díaz-Ferguson, Hill, & Shea, 2014; Pilliod, Goldberg, Arkle, & Waits, 2013; Thomsen, Kielgast, Iversen, Wiuf et al., 2012), in natural freshwater systems (e.g., Doi et al., 2017; Lacoursière-Roussel, Côté, Leclerc, & Bernatchez, 2016) and in marine environments (Jo et al., 2017; Yamamoto et al., 2016) demonstrate that eDNA can be used to measure relative population abundance with a species-specific primer set and qPCR. While many more controlled experiments are needed in all ecosystems to determine the relationship of abundance to copy number observed in qPCR, evidence thus far from water samples signifies that eDNA contains information about a species' relative abundance.

Overall, ascertaining abundance information using metabarcoding of eDNA for whole communities still lacks substantial evidence, but some studies in aquatic environments have shown positive relationships between the relative number of reads and relative or rank abundance estimated with conventional methods. Evans et al. (2016) showed in a mesocosm setting that relative abundance of individuals and biomass was correlated with relative read abundance in mesocosms containing fishes and an amphibian. In a natural lake, Hänfling et al. (2016) found that the rank abundance derived from long-term monitoring was correlated with read abundance for fish species, and positively correlated with gillnet surveys conducted at the same time as eDNA sampling. In deep sea habitats, Thomsen et al. (2016) found that when reads for fish were pooled to the taxonomic rank of families, there was a correlation with relative abundance of individuals and biomass captured in trawls. While these examples are promising, not all studies support such findings (e.g., Lim et al., 2016).

Challenges to accurate abundance estimation through eDNA metabarcoding stem from multiple factors in the field and the laboratory (Kelly, 2016). In the field, the copy number of DNA arising from an individual in an environmental sample is influenced by the characteristics of the "ecology of eDNA" (e.g., its origin, state, fate and transport; Barnes & Turner, 2016). Because different animal and plant species are likely to have different rates of eDNA production or "origin" (Klymus, Richter, Chapman, & Paukert, 2015), exhibit different "transport" rates from other locations (Civade et al., 2016; Deiner & Altermatt, 2014) or stability or "fate" of eDNA in time (Bista et al., 2017; Yoccoz et al., 2012), eDNA in an environmental sample could be inconsistent relative to a species' true local and current abundance. Therefore, continued research on how the origin, state, fate and transport of eDNA influence estimates of relative abundance is needed before we can understand the error this may generate in our ability to estimate abundance.

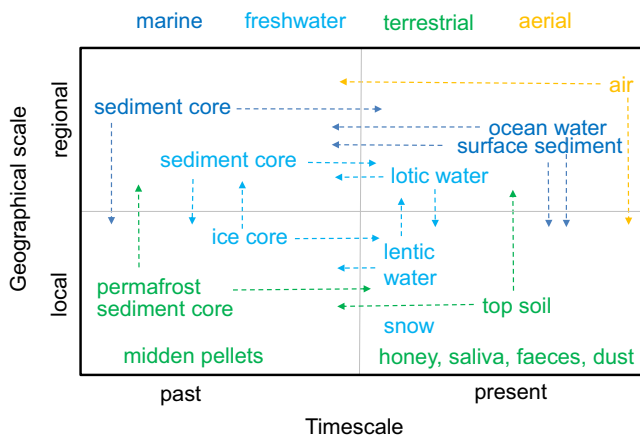
In the laboratory, primer bias driven by mismatches with their target has been shown to skew the relative abundance of amplified DNA from mock communities (Elbrecht & Leese, 2015; Piñol, Mir, Gomez-Polo, & Agustí, 2015). Similarly, the same mechanism could

**TABLE 1** Representative studies comparing richness estimates with traditional sampling or historical data for a geographic location to that of eDNA metabarcoding

Habitat	Macro-organism taxonomic focus	eDNA sample type	Traditional sampling method	eDNA efficacy finding <sup>a</sup>	Authors, Year
Air	Plants	Air pollen trap	Morphological identification	Better taxonomic resolution	Kraaijeveld et al. (2015)
Freshwater	Fish	Flowing water	Depletion-based electrofishing	Higher diversity	Olds et al. (2016)
Freshwater	Invertebrates	Flowing water	Kicknet in stream and historical data	Higher diversity	Deiner et al. (2016)
Freshwater	Fish	Stagnant water	Gillnet, trapping, hydroacoustics, analysis of recreational anglers' catches	Complementary	Hänfling et al. (2016)
Freshwater	Reptiles, amphibians	Stagnant water	Species distribution model based on historical data (i.e., distribution range and habitat type)	Increase species distribution knowledge	Lacoursière-Roussel, Dubois, Normandeau & Bernatchez, (2016)
Freshwater	Amphibians, fish	Stagnant water; flowing water	Amphibians: visual encounter survey, mesh hand-net; fish: electrofishing, and/or netting protocols (fyke, seine, gill)	Greater detection probability	Valentini et al. (2016)
Freshwater	Amphibians, fish, mammals, invertebrates	Stagnant water; flowing water	Active dip-netting, fresh tracks or scat, electrofishing with active dip-netting	Complementary	Thomsen, Kiekgast, Iversen, Wiuf et al. (2012)
Freshwater	Fish	Stagnant water; flowing water; surface sediment	Fyke net	Higher diversity	Shaw et al. (2016)
Freshwater	Invertebrates	Water column; surface sediment	Sediment collected using a Van Veen grab	Higher diversity	Gardham, Hose, Stephenson, and Chariton (2014)
Freshwater	Fish/Diptera	Surface and bottom water column	Long-term data, electrofishing (fish) and emerging traps (Diptera) at the time of eDNA sampling	Higher diversity compared to sampling but lower diversity compared to long-term data	Lim et al. (2016)
Marine	Fish	Surface and bottom water column	Long-term observation	Complementary	Yamamoto et al. (2017)
Marine	Fish	Bottom water column	Trawl catch data	Similar family richness	Thomsen et al. (2016)
Marine	Fish	Water column	Scuba diving	Higher diversity	Port et al. (2015)
Terrestrial	Plants	Honey	Melissopalynology (i.e., pollen grains retrieved from honey are identified morphologically)	Complementary	Hawkins et al. (2015)
Terrestrial	Mammals, plants	Midden pellets	Historical surveys	Higher diversity	Murray et al. (2012)
Terrestrial	Mammals	Saliva	Local knowledge (i.e., physical evidence) and camera data	Complementary	Hopken et al. (2016)
Terrestrial	Birds, invertebrates, plants	Top soil	Invertebrates: leaf litter samples and pitfall traps; reptiles: pitfall traps and under artificial ground covers; birds: distance sampling method; plants: above-ground surveys	Complementary for plants and invertebrates	Drummond et al. (2015)
Terrestrial	Earthworms	Top soil	Irrigated quadrats with 10 L of allyl isothiocyanate solution and hand-collected emerging worms	Complementary	Pansu et al. (2015)
Terrestrial	Plants	Top soil	Historical surveys	Complementary	Jørgensen et al. (2012)
Terrestrial	Plants	Top soil	Above-ground surveys	Complementary and better taxonomic resolution	Yoccoz et al. (2012)
Terrestrial	Vertebrates	Top soil	Local knowledge from safari parks, zoological gardens and farms; visual observations; historical surveys	Complementary	Andersen et al. (2012)

<sup>a</sup>Complementary means the two survey methods detected different diversity, but does not exclude that some of the diversity was detected by both methods. Higher diversity means the study found more diversity was detected compared to conventional, but does not exclude that some of the diversity was *not* detected by both methods. Better taxonomic resolution means that sequence-based identifications could be resolved to a lower taxonomic rank compared with the conventional method.



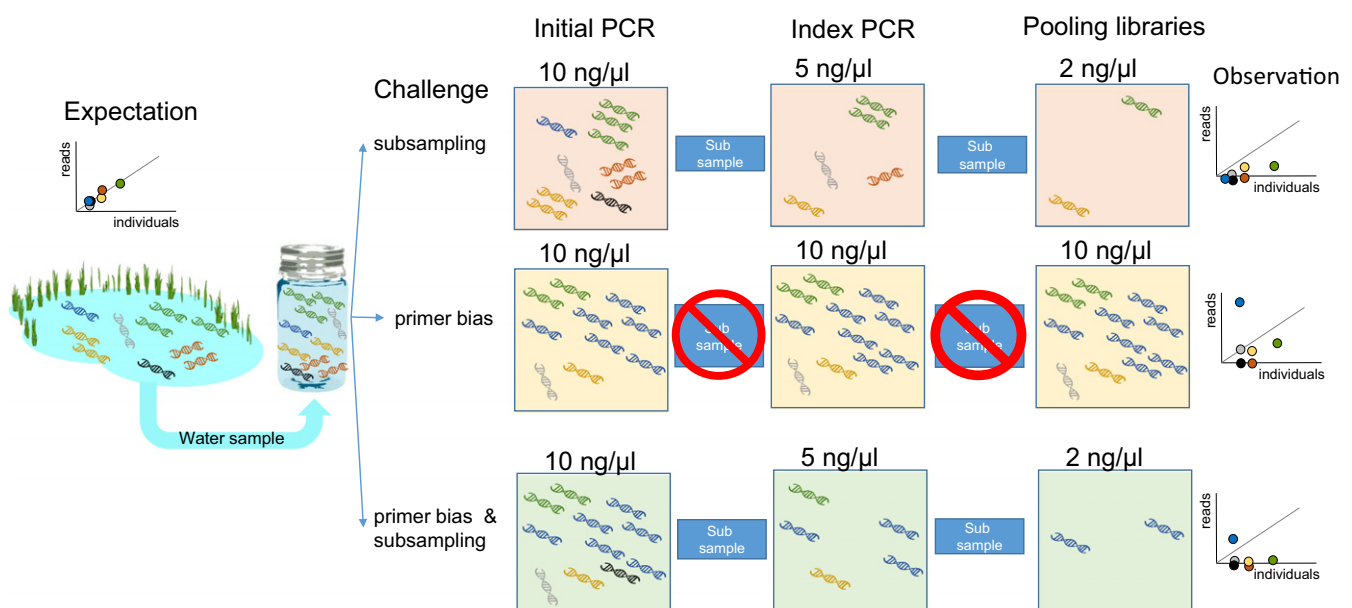


**FIGURE 1** Environmental DNA sample types have different spatial and temporal scopes of inference from different habitats. Consider each sample type as a single sample from that environment. Placement of a sample type in a quadrant is not quantitative, but represents a common scale at which it has been used. Dashed arrows indicate the potential for a sample type to confer information at multiple scales of inference, but additional research to quantify these possibilities is needed

alter the relative abundance of a species' DNA amplified from eDNA (Figure 2). Primer bias results in an increased variance in abundance of reads observed relative to their true abundance in an environmental sample (Figure 2). Another source of error is related to library preparation methods. Analysis of mock communities has shown that amount of subsampling during processing steps can drive the loss of rare reads (Leray & Knowlton, 2017) and likely occurs for eDNA samples as well (Shelton et al., 2016). The combination of primer

bias and library preparation procedures alone could cause a large variance in reads observed for any given species and could prevent rare species detection altogether (Figure 2). Technical approaches and potential solutions to alleviate primer bias and alternative library preparation methods are discussed in the "Challenges in the field, in the laboratory and at the keyboard" section. While in the end, it may be that eDNA metabarcoding is not the most accurate method for simultaneously measuring the relative abundance for multiple species from eDNA, researchers should consider whether the eDNA metabarcoding method is accurate enough for application in a particular study or an applied setting. Other methods such as capture enrichment are being examined and are promising because they avoid PCR and hence the bias this may cause, but they do require extensive knowledge of the biodiversity to design targeted gene capture probes and they come with a greater costs for analysis (Dowle, Pochon, Banks, Shearer, & Wood, 2016). Future studies comparing qPCR, eDNA metabarcoding and capture enrichment will be beneficial to determine which method yields accurate estimates of relative abundance from eDNA.

Before ruling out the plausibility entirely, in the short term, simulations could certainly be used to test the effects of technical laboratory issues and account for the ecology of eDNA to decipher under what conditions reliable estimates for abundance can be achieved from eDNA metabarcoding. Promising steps in this direction have been investigated through simulation to learn the nature of how data sets deliberately "noised" conform to neutral theory parameters in estimation of rank abundance curves (Sommeria-Klein, Zinger, Taberlet, Coissac, & Chave, 2016). Results from simulations studies such as this could then be used to inform mock community



**FIGURE 2** Challenges for estimating abundance from environmental DNA metabarcoding. For simplicity, assume one DNA molecule depicted in the pond is equal to one organism and colours represent different species. Additionally for this example, assume that sampling is not biased (i.e., DNA copies are sampled in their true abundance), that boxes surrounding DNA molecules represent 1  $\mu$ l and that one DNA molecule represents 1 ng of DNA. Thus, values illustrated show the effect of primer bias, subsampling and their combination on the ability to estimate abundance

experiments and test hypotheses (e.g., type of error distribution expected) under realistic seminatural environments.

### 3 | ECOSYSTEMS, THEIR SAMPLE TYPES AND KNOWN SCALES OF INFERENCE IN SPACE AND TIME

#### 3.1 | Freshwater ecosystems

Environmental DNA metabarcoding of different sample types has been highly successful in obtaining species richness estimates for animals in aquatic systems (Figure 1, Table 1). In one of the first seminal studies, Thomsen et al. (2012a) used surface water from lakes, ponds and streams in Denmark to demonstrate that eDNA contained information about aquatic vertebrate and invertebrate species known from the region. However, there is a notable lack of eDNA metabarcoding studies assessing living aquatic plant communities, and this remains an open area for further research.

Mounting evidence suggests that the spatial and temporal scale of inference for eDNA sampled from surface water differs for rivers and lakes (Figure 1). Specifically, river waters measure species richness present at a larger spatial scale (Deiner et al., 2016) compared to eDNA in lake surface waters (Hänfling et al., 2016). Differences between lake and river eDNA signals may be due to the transport of eDNA over larger distances in rivers compared to longer retention times of water in lake systems (Turner, Uy, & Everhart, 2015). However, lakes and ponds with river and surface run-off inputs, combined with lake mixing or stratification, may serve as eDNA sources for catchment-level terrestrial and aquatic diversity estimates similar to rivers (Deiner et al., 2016). No studies to date have estimated the sources of eDNA in surface water from a lake's catchment and related it to the diversity locally occurring in the lake. However, ancient DNA from sediment cores in lakes (sedaDNA) has been used to determine historical plant (e.g., Pansu, Giguët-Covex, Ficetola, et al. 2015; Parducci et al., 2013) and livestock communities (Giguët-Covex et al., 2014), thus indicating that lakes do receive DNA from species in their catchments which can be incorporated into their sediments. For a more extensive review of sedaDNA being used to reconstruct past ecosystems, see Pedersen et al. (2015) and Brown and Blois (2016).

Most often, species richness estimates generated from eDNA in surface waters of lakes and rivers reflect recent site biodiversity, while those from eDNA found in surface sediments may reflect a temporally extended accumulation of eDNA. For example, Shaw et al. (2016) compared estimates of fish species richness from water and surface sediment samples. Generally, they found species were detected in both samples, but estimates of species richness from water samples were in better agreement with the species physically present at the time of sampling. The temporal scale of inference in surface sediments is largely unknown and needs further examination (Figure 1).

In addition to surface freshwater (~1%), groundwater (~30%) and ice (~69%) comprise much of earth's freshwater (Gleick, 1993).

While the other freshwater habitats far surpass the amount of surface water, their extant biodiversity is rather poorly described (Danielopol, Pospisil, & Rouch, 2000). Groundwater is known to harbour a wide range of specialist taxa which are difficult to assess using conventional survey methods due to the inaccessibility of these habitats (Danielopol et al., 2000). Groundwater microorganism metabarcoding studies have shown high fungal (Sohlberg et al., 2015) and bacterial (Kao et al., 2016) diversity, and there are examples of species-specific studies on the cave-dwelling amphibian *Proteus anguinus* (e.g., Goricki et al., 2017; Vörös, Márton, Schmidt, Gál, & Jelić, 2017). However, there is a clear lack of eDNA metabarcoding studies that could shed light on the diversity of a wide range of macro-organisms known to inhabit groundwater, including turbellarians, gastropods, isopods, amphipods, decapods, fishes and salamanders. The spatiotemporal scale of inference of eDNA samples from groundwater is currently unknown. Surveying eDNA in systems with knowledge of the complex hydrology and interactions between surface and groundwater will be interesting places to start to reveal the scale of inference for eDNA surveys for these environments.

Environmental DNA found in sediment cores and ice core sediments generally reflects a historical biodiversity sample (Figure 1) and is more commonly used as a source of ancient DNA (Willerslev et al., 2007). To date, animal and plants surveyed from lake sediment cores suggest that information about terrestrial and aquatic communities can be estimated as far back as 6 to 12.6 thousand years before present (Giguët-Covex et al., 2014; Pedersen et al., 2016), whereas eDNA from sediments in ice cores have successfully been used to reconstruct communities as far back as 2000 years before present (Willerslev, Hansen, Christensen, Steffensen, & Arctander, 1999). The spatial scale of inference for sediment samples types has not been tested, but when samples from multiple locations are combined, large areas can be surveyed for the past presence of species (Anderson-Carpenter et al., 2011). For modern communities, snow has served as a viable sample type and enabled a local survey of wild canids in France (Valiere & Taberlet, 2000). Environmental DNA metabarcoding of water from glacial run-off will also likely be a valuable tool to survey animal and plant richness living in glacial and subglacial habitats, which are undergoing dramatic change because of climate warming (Giersch, Hotaling, Kovach, Jones, & Muhlfeld, 2017).

#### 3.2 | Marine ecosystems

The use of eDNA metabarcoding is often described as challenging in marine ecosystems, due to the potential dilution of eDNA in large volumes of water and additional abiotic factors (salinity, tides, currents) that likely impact eDNA transport and degradation (Foote et al., 2012; Port et al., 2016; Thomsen et al., 2012b), not to mention the logistics involved in undertaking such surveys. Nonetheless, eDNA metabarcoding surveys of marine fish from coastal water samples have demonstrated that eDNA can detect a greater taxonomic diversity compared to conventional survey techniques (Table 1), while simultaneously improving the detection of rare and vagrant



fish species and revealing cryptic species otherwise overlooked by visual assessments (O'Donnell et al., 2017; Port et al., 2016; Thomsen et al., 2012b, 2016). Marine mammals have been surveyed with acoustic surveys and eDNA metabarcoding, and here, the conventional acoustic methods detected a greater species richness (Foote et al., 2012). Nevertheless, this study used low sample volumes compared to other marine studies (15–45 ml vs. 1.5–3.0 L), and the authors concluded that larger sample volumes would likely lead to greater similarity between eDNA and conventional methods. In Monterey Bay, California, water sampled from depths <200 m or >200 m were used to detect marine mammals such as seals, dolphins and whales in addition to many fishes and sharks (Andruszkiewicz et al., 2017). The taxonomic groups detected were spatially explicit and were found more or less in water associated with their expected habitat.

Longitudinal transport of animal and plant eDNA in marine environments is not well studied. But, similar to freshwater sediment cores from lakes, vertical transport into marine sediments is likely to preserve a large proportion of eDNA from particulate organic matter or eDNA that has become directly adsorbed onto sediment particles. This absorption shields nucleotides from degradation (particularly oxidation and hydrolysis) and facilitates long-term preservation of genetic signals over potentially large spatiotemporal scales (Figure 1). Marine sediment eDNA concentrations have been shown to be three orders of magnitude higher than in seawater eDNA (Torti, Lever, & Jørgensen, 2015) and eDNA from both ancient and extant communities is typically recovered (Lejzerowicz et al., 2013). Similar to lake sediments, marine sediments can accumulate genetic information from both terrestrial and pelagic sources (Torti et al., 2015).

Marine sediments are difficult to sample because of the logistical effort involved in obtaining samples, which often requires ship time and specialized coring equipment. Even though much work remains to be carried out to understand the spatiotemporal scale of inference for marine sediment cores, comparisons between eDNA and environmental RNA (eRNA) metabarcoding are hypothesized to allow inference between present and past diversity. Environmental RNA is thought to be only available from live organisms in the community; thus, the comparison between eDNA and eRNA has been investigated. In applied settings, eDNA metabarcoding of surface sediments has revealed benthic impacts of aquaculture for Atlantic salmon farming on short spatial scales using both eDNA and eRNA (Pawłowski, Esling, Lejzerowicz, Cedhagen, & Wilding, 2014). Guardiola et al. (2016) showed through a comparison of eDNA and eRNA that spatial trends in species richness from these two sources were similar, but that eDNA detected higher diversity. Overall, the fate, transport and decomposition of animal and plant eDNA in marine environments are poorly known compared to other environments, and there is pressing need for further studies.

### 3.3 | Terrestrial and aerial ecosystems

Environmental DNA from terrestrial sediment cores is a valuable tool for investigating past environments and reconstructing animal and

plant communities (Figure 1; Haouchar et al., 2014; Jørgensen et al., 2012; Willerslev et al., 2003). Animal remains also provide opportunities to reconstruct past trophic relationships. For example, eDNA metabarcoding of pellets in herbivore middens has been used to identify species in ancient animal and plant communities (Figure 2; Murray et al., 2012) and DNA traces from microplant fossils within coprolites were used to reconstruct former feeding relationships in rare and extinct birds (Wood et al., 2012). Again here, the recent reviews of Brown and Blois (2016) and Pedersen et al. (2015) provide a more extensive overview for how ancient DNA is used to uncover past animal and plant communities.

In modern environments, eDNA isolated from top soils has been used to characterize biodiversity in earthworms (Bienert et al., 2012; Pansu, De Danieli, Puissant, et al. 2015), invertebrates (McGee & Eaton, 2015), plants (Taberlet, Prud'Homme, Campione, et al. 2012; Yoccoz et al., 2012) and vertebrate species (Andersen et al., 2012). In what is perhaps the most comprehensive analysis using eDNA metabarcoding for any environment, Drummond et al. (2015) simultaneously surveyed all three domains of life in top soil using PCR primers that amplified five different metabarcoding regions, thus demonstrating the power of this method for assessing total richness for an area. However, the spatial scale of inference for many terrestrial eDNA samples is an open question (Figure 1). Research on the timescale of inference for DNA in top soil suggests that long fragments of DNA break down quickly, but short fragments remain detectable for days to years after the presence of the species (Taberlet, Prud'Homme, Campione, et al. 2012; Yoccoz et al., 2012). Thus, the fragment length amplified can change the temporal resolution of a soil sample.

There are many additional sources for eDNA sampling besides soil in terrestrial ecosystems. For animals, bloodmeals from leeches (Schnell et al., 2012) and carrion flies (Calvignac-Spencer et al., 2013) have been used to survey mammal diversity. Saliva on browsed twigs was tested as a source of eDNA to survey ungulates (Nichols, Koenigsson, Danell, & Spong, 2012) and on predated eggs and carcasses of ground-nesting birds to discover predators or scavengers (Hopken, Orning, Young, & Piaggio, 2016). DNA extracted from spider webs has also been used to detect spiders and their prey (Xu, Yen, Bowman, & Turner, 2015). For plants, pollen within honey has revealed honey bee foraging preferences (De Vere et al., 2017; Hawkins et al., 2015). Craine et al. (2017) surveyed dust from indoor and outdoor environments throughout the United States and found that plant DNA from known allergens was almost twice as high outdoor compared with indoor environments. In addition to allergen detection from pollen, there remain many potential applications of dust eDNA to assess animal species richness. Faecal DNA has also been used as a source of eDNA to assess diet composition, but most studies utilizing this source of eDNA are focused on single-species detections and population genetic inferences (see review from Rodgers & Janečka, 2013) and are not necessarily using eDNA sources from faecal DNA to estimate species richness of terrestrial communities. Boyer, Cruickshank, and Wratten (2015) proposed that surveys of faeces from generalist predators can act as “biodiversity capsules” and analysis of this eDNA source should give rise to biodiversity surveys for prey communities

in landscapes. While all of these sources are available, most of these sample types (e.g., leaves from a tree, faecal pellets, spider webs and dust) do not have a known scale of inference in space and time. A single sample of eDNA from these sources is not likely to confirm species richness for more than a local scale, but combination of multiple sample sources (e.g., leaves, faecal pellets and spider webs throughout a park) sampled over time may allow for spatial and temporal estimates of terrestrial species richness.

Surveys of airborne eDNA have placed greater emphasis on the detection of bioaerosols that cause infection or allergic responses in animals and plants (West, Atkins, Emberlin, & Fitt, 2008). For example, Kraaijeveld et al. (2015) investigated airborne pollen that can cause hay fever and asthma in humans and showed that the source of allergenic plant pollen could be identified more accurately using eDNA from plant pollen filtered from the air compared to microscopic identification. A particularly interesting area for further research is to gain an understanding of the scale of inference for air samples in space and time (Figure 1). While plant eDNA can be ascertained, surveying other species such as birds

and insects from aerial eDNA sources has not been tested to our knowledge.

#### 4 | CHALLENGES IN THE FIELD, IN THE LABORATORY AND AT THE KEYBOARD

Despite the obvious power of the approach, eDNA metabarcoding is affected by a host of precision and accuracy challenges distributed throughout the workflow in the field, in the laboratory and at the keyboard (Thomsen & Willerslev, 2015). Following study design (e.g., hypothesis/question, targeted taxonomic group, Figure 3), the current eDNA workflow consists of three components: field, laboratory and bioinformatics. The field component consists of sample collection (e.g., water, sediment, air) that is preserved or frozen prior to DNA extraction. The laboratory component has four basic steps: (i) DNA is concentrated (if not performed in the field) and purified, (ii) PCR is used to amplify a target gene or region, (iii) unique nucleotide sequences called "indexes" (also referred to as "barcodes") are



**FIGURE 3** Important guiding questions for consideration in the design and implementation phases of an environmental DNA metabarcoding study

incorporated using PCR or are ligated onto different PCR products, creating a “library” whereby multiple samples can be pooled together, and (iv) pooled libraries are then sequenced on a high-throughput machine (most often the Illumina HiSeq or MiSeq platform). The final step after laboratory processing of samples is to computationally process the output files from the sequencer using a robust bioinformatics pipeline (Figure 3, Box 2). Below, we emphasize the important and rapidly evolving aspects of the eDNA metabarcoding workflow and give recommendations for ways to reduce error.

#### 4.1 | In the field

As for any field study, the study design is of paramount importance (Figure 3, Box 2), as it will impact the downstream statistical power and analytical interpretation of any eDNA metabarcoding data set. For example, sampling effort and replication (especially biological) are positively correlated with the probability of detecting the target taxa (Furlan, Gleeson, Hardy, & Duncan, 2016; Willoughby, Wijayawardena, Sundaram, Swihart, & DeWoody, 2016). Despite the extensive evidence of the occurrence of macro-organism DNA in the environment, our fundamental understanding of what “eDNA” is from any environmental sample is still lacking. For an illustration of this challenge, we summarize what is known about eDNA in freshwater environments. The current state of the art relies on the fact that we can access eDNA by precipitating DNA from small volumes of water samples (e.g., 15 ml, Ficetola, Miaud, Pompanon, & Taberlet, 2008), or filter eDNA from the water column using a variety of filter sizes (0.22  $\mu\text{m}$  and upwards; Rees, Maddison, Middleditch, Patmore, & Gough, 2014). Filtration protocols lead to a working hypothesis that aqueous eDNA is either derived from cellular or organellar sources (e.g., mitochondria, Lacoursiere-Roussel, Rosabal, & Bernatchez, 2016; Turner et al., 2014; Wilcox, McKelvey, Young, Lowe, & Schwartz, 2015), and precipitation protocols suggest extracellular sources (Torti et al., 2015). It is clear that at least some freshwater eDNA comes from intact cellular or organellar sources because it has recently been demonstrated to be available in the genomic state (Deiner, Renshaw, et al. 2017). Thus, eDNA in water exists in both undegraded and degraded forms (Deiner, Renshaw, et al. 2017). However, continued research on the origin, state and fate of eDNA will greatly inform numerous strategies regarding its acquisition (filtering, replication, sample volumes and spatial sampling strategies; Barnes & Turner, 2016). Many methods for solving current challenges of false negatives (e.g., use of biological replicate sampling, improved laboratory methods) and false positives (e.g., use of negative controls) in the field are explored in a recent review (Goldberg et al., 2016); we therefore refer readers to this review rather than treat those topics in-depth here.

#### 4.2 | In the laboratory

There are a number of recent studies that focus on the capture, preservation and extraction of eDNA, and the literature reviewed

therein summarizes the important considerations and trade-offs that should be tested before a large-scale study is conducted (e.g., Deiner, Walser, Mächler, & Altermatt, 2015; Renshaw, Olds, Jerde, McVeigh, & Lodge, 2015; Spens et al., 2017). Rather than reiterate those aspects here, we focus on primer choice and library preparation. For animal and plant studies, PCR primers most often target mitochondrial or plastid loci or nuclear ribosomal RNA genes (Table S2). The standard barcoding markers defined by the Consortium for the Barcode of Life (CBOL) are cytochrome c oxidase subunit I (COI or *cox1*), for taxonomical identification of animals (Hebert, Ratnasingham, & de Waard, 2003), and a two-loci combination of *rbcL* and *matK* as the plant barcode (Hollingsworth et al., 2009) with ITS2 also suggested as valid plant barcode marker (Chen et al., 2010). However, there are limitations for using the standard barcoding markers in macro-organism eDNA metabarcoding. Specific to COI, other DNA regions are commonly used because not all taxonomic groups can be differentiated to species equally well (Deagle, Jarman, Coissac, Pompanon & Taberlet, 2014) and because it is challenging to design primers in this gene for a length that is suitable for short amplicon analysis, but some regions have been identified (Leray et al., 2013). The most common alternative markers used are mitochondrial ribosomal genes such as 12S and 16S or protein-coding genes such as cytochrome B (Table S2). Specific to the plant barcoding loci, the two-loci combination primarily used for barcoding plants can be independently generated, but is not always possible to recover which fragment from each gene is associated with each other in an eDNA sample, rendering species identification using the standard plant barcode challenging. Bioinformatic methods can help resolve these situations to some extent, and may work when diversity is low in a sample (Bell et al., 2016). Therefore, often one or different markers are used (e.g., P6 loop of the *trnL* intron; Sønsteby et al., 2010; Taberlet et al., 2007; Table S2).

Additionally, some highly evolving noncoding loci, such as ITS rRNA, are used (Table S2), but these markers do not always allow for the construction of alignments to determine MOTUs during data analysis because they have intragenomic variation that complicates their use in biodiversity studies (plant ITS rRNA may be an exception; Bell et al., 2016). For these loci, an unknown environmental sequence is often discarded unless it has an exact database match reducing a data set to only known and sequenced biodiversity. Due to these factors, other metabarcoding loci such as 18S rRNA genes may be more appropriate (e.g., in studies of marine invertebrates, Bik et al., 2012), especially if phylogenetic analysis is needed to narrow down taxonomic assignments and circumvent database limitations (Box 3).

Once the locus or loci are chosen, primers are then designed based on the taxonomic group(s) of interest within a study, and the need for broad (multiple phyla) vs. narrow (single order) coverage to test study-specific hypotheses (Figure 3). When choosing previously designed primers (Table S2) or when designing new primers, it is important to perform rigorous testing, *in silico*, *in vitro* and *in situ* to infer their utility for metabarcoding eDNA in a new study system (Elbrecht & Leese, 2017; Freeland, 2016; Goldberg et al., 2016).

**BOX 2 BASIC BIOINFORMATIC PIPELINE FOR EDNA METABARCODING FOR PLANTS AND ANIMALS**

Bioinformatic processing of sequence data is one of the most critical aspects of eDNA metabarcoding studies, helping to substantiate research findings, following field and laboratory work components. Standardization of bioinformatics in a "pipeline" can ensure quality and reproducibility of findings; however, some level of customization is required across studies. Customization is needed to compensate for advances in sequencing technology, software workflows and the question being addressed. Therefore, taking raw read data and turning it into a list of taxa require multiple quality assurance steps—some necessary, while others optional. Reaching an absolute consensus for the approaches and software used is not necessary as these will always be in flux, but here we advise careful consideration of the following preprocessing steps *at a minimum* for HTS data before embarking on further analyses (e.g., for biodiversity estimates and statistical significance). We focus primarily on processing Illumina generated data sets, and therefore if the technology is different, many of the bioinformatic tools highlighted and advice is transferable to preprocessing of data produced on other platforms, but may be different.

**Terms****Chimeras**

PCR artefacts made of two or more combined sequences during the extension step of PCR amplification.

**Phred quality score**

Quality scoring per nucleotide for Illumina sequencing providing the probability that a base call is incorrect.

**Sequence merging**

Combining forward (R1) and reverse (R2) reads from paired-end (PE) sequencing, using criteria such as minimum overlap or quality score.

**Sequence trimming**

The process of cutting/removing the beginning or end of sequencing reads can be performed either by searching for a specific sequence (removal of adaptors, indexes and primers) or based on quality score.

**Singletons**

MOTUs that appear only once in the data are likely to be rare taxa, false positives, low-level contamination or unremoved chimeras, and should be treated with appropriate consideration.

**Primer-adaptor trimming**

Preliminary steps of bioinformatics processing include demultiplexing of the samples based on the indices used (unique nucleotide tags incorporated into raw sequence data) and trimming (i.e., removal) of the adaptor sequences. The adaptors are specific DNA fragments that are added during library preparation for ligation of the DNA strands to the flow cell during Illumina sequencing. Additionally, the index sequences themselves and the primer sequences should be trimmed (e.g., using software such as Cutadapt, Trimmomatic, QIIME), allowing either zero or a low level of mismatch between the exact sequence of the primer or index and the observed reads.

**Merging or end trimming**

Sequences from Illumina runs tend to drop in quality towards the 3' end of the reads, as phasing leads to increased noise (and lower signal) in later chemistry cycles. Thus, the quality score of reads should be reviewed to allow informed decisions on the appropriate length of end trimming (single-end runs), merging (paired-end runs) and subsequent sequence quality filters. Visualizing the quality scores from raw reads or demultiplexed sequences (using software such as FastQC) will help with the selection of downstream quality cut-off levels.

When paired-end (PE) sequencing is used for an amplicon of suitable size, the forward (R1) and reverse (R2) reads should be combined (merged) to form the complete amplicon. Using merged sequences improves accuracy as the lower-quality bases at the tail ends of individual reads can be corrected based on the combined reads. Here, the minimum overlap for R1 and R2 reads should be specified and "orphan" reads with little or no overlap between forward and reverse pairs can be discarded. Inspection of the quality scores, as mentioned above, can provide an estimate of optimal parameters for merging of R1 and R2 reads. Even though a specific consensus does not exist yet; in many cases, an overlap of at least > 20 bp is selected (Deiner et al., 2015; Gibson et al., 2015).

**Quality filtering**

For most HTS platforms, a Phred score is calculated and subsequently used to determine the maximum error probabilities (Bokulich et al., 2013). Selected strategies include filtering based on a lower Phred score cut-off, usually set at least above 20 or 30 (Bista

(Continues)

**BOX 2** (Continued)

et al., 2017; Elbrecht & Leese, 2015; Hänfling et al., 2016). Quality filtering can also be performed based on maximum error (maxee) probability, which is also derived from Phred scores. The lower the maximum error, the stricter the cut-off. Selection of a maximum error filtering level of 1 or 0.5 is common in macro-organism studies (Bista et al., 2017; Pawlowski et al., 2014; Port et al., 2016). Additionally, in the case of single-end sequencing, or when long amplicons without sufficient overlap of the forward and reverse reads are used, it is advised that trimming should be performed from the appropriate end. It is often the case that reads are trimmed to a common length, which facilitates alignment downstream and minimizes miscalled bases as a merging step cannot be used.

**Removing short reads**

Many studies also select to remove short reads from the data set before clustering as the presence of high length variation could influence the clustering process (see USEARCH manual, Edgar, 2010). These sequences could result from sequencing of primer dimers that have not been removed (Pawlowski et al., 2014). Different studies select a variety of minimum length reads, from very short 20 bp (Valentini et al., 2016), to medium 60–80 bp (Pawlowski et al., 2014; Shaw et al., 2016) and up to 100 bp (Bista et al., 2017; Gibson et al., 2015; Hänfling et al., 2016; Pawlowski et al., 2014). Note that some demultiplexing or quality filtering workflows may automatically set a minimum sequence length when processing input data and it is advisable to check whether such a parameter is included by default.

**Removing singletons and chimeras**

Important steps after MOTU clustering involve removal of singletons and chimeras. Chimeras are by-products of the PCR amplification process from two or more parental sequences (chimeric), most commonly produced through an incomplete extension step (Edgar, Haas, Clemente, Quince, & Knight, 2011). It has been shown that when unique reads, such as chimeras and singletons, are withheld in analysis, the estimation of diversity can be severely inflated (Kunin, Engelbrektson, Ochman, Hugenholtz, 2010). The nature of the chimeric sequences, which can be present as high-quality reads, does not enable their removal directly through quality-based end trimming (Coissac et al., 2012). Removal of chimeras can be performed either de novo or based on a reference database. Most common practice to date is the de novo method as a sufficient reference database may not be available. Despite the variation in software used such as UCHIME (Edgar et al., 2011), OBITOOLS (Boyer et al., 2016) or CHIMERASLAYER (Haas et al., 2011), there is a consensus regarding the importance of removing chimeras and singletons as a minimum quality control for bioinformatics pipeline.

**Abundance filtering**

In addition to quality filtering based on quality scores and removal of chimeras and singletons, many studies also employ further filtering for removal of low abundance sequences (Murray et al., 2015). This step arises from the need to control for laboratory contamination or because of cluster contamination on the flow cell (unique to Illumina platforms; Olds et al., 2016).

The process of applying abundance filtering requires setting an MOTU abundance threshold by which MOTUs are only retained in analysis if their relative abundance is higher than the selected threshold (Bokulich et al., 2013). Selection of a threshold varies between studies and there is no generally accepted definition of what constitutes an insufficiently abundant read (Murray et al., 2015), perhaps with the exception of singletons. Abundance filtering may be applied minimally or avoided entirely, especially if stringent quality trimming parameters are applied to raw reads and detection of “rare” MOTUs is an important aspect of a study (Bokulich et al., 2013). Another option that could be used involves selection of a threshold based on availability of empirical data as was performed in Valentini et al. (2016). An increasing number of studies have employed the sequencing of positive controls to establish a threshold level (Hänfling et al., 2016; Port et al., 2016; Stoeckle, Soboleva, & Charlop-Powers, 2017). Technical replicates can also be used to assess consistency as was shown to be effective with assessing omnivore diets (De Barba et al., 2014).

Using a positive control-defined error level works by identifying the abundance of sequences in the control sample that belong to nontarget taxa and can be the result of errors such as contamination. Furthermore, the distribution of *phiX* reads assigned to target samples has been used to investigate the presence of “tag-jumps” (Schnell, Bohmann, et al. 2015) and mis-assigned reads during demultiplexing (Hänfling et al., 2016; Olds et al., 2016). The exact mechanisms for mis-assignment of reads remain unknown, but increasingly many studies are reporting this error to be between 0.01% and 0.03% of reads (Hänfling et al., 2016; Olds et al., 2016; Stoeckle et al., 2017). Adjustments for this include use of a threshold approach-based on negative and/or positive controls and remove a low number of reads from any given sample. The issue of abundance filtering most significantly causes uncertainty in low abundance MOTUs and will continue to be a problem for the detection of rare species. Therefore, to avoid negative impacts to scientific insights or management decisions, careful consideration and transparency regarding how technical artefacts are dealt with during bioinformatic data analysis is needed until these artefacts are well understood.

(Continues)



**BOX 2** (Continued)**Recording removed data**

For all quality control steps, the data removal should be transparent. Often studies report the total number of sequences obtained, but then rarely show how each quality filtering step affects the number of sequences used in testing ecological hypothesis nor do researchers provide the subset of sequences that were retained or omitted. Deleting data without a clear justification does not allow transparency. Therefore, including a supplemental table in eDNA metabarcoding studies showing the number of sequences remaining after each filtration step is advised and archiving the subset of reads retained after each filtering step on a platform such as DRYAD (<http://datadryad.org/>) or archiving the exact pipeline with version control information on a platform such as GITHUB (<https://github.com/>) will allow for greater transparency and reproducibility of quality filtering.

Amplicon size is also an important consideration because there may be a trade-off in detection with amplicon length (e.g., short fragments are more likely to amplify). However, short fragments may persist longer in the environment and increase the inference in space or time that can be made from an environmental sample (Bista et al., 2017; Deagle, Eveson, & Jarman, 2006; Jo et al., 2017; Yoccoz et al., 2012). Additionally, use of more than one locus for a target group can allow for tests of consistency between loci and increase stringency of detection for any species (Evans et al., 2017).

Once primers are designed and PCR products are amplified, eDNA metabarcoding relies on multiplexing large numbers of samples on HTS platforms in order to make the tool cost-effective. Illumina (MiSeq and HiSeq) sequencing platforms at the moment outperform other models for accuracy (Loman et al., 2012), and multiplexing samples is usually achieved by the incorporation of sample-specific nucleotide indices and sequencing adapters during PCR amplification. However, multiplexing creates opportunities for errors and biases. In this facet of the workflow, it is important to avoid methods that induce sample-specific biases in amplification (O'Donnell, Kelly, Lowell, & Port, 2016) and to reduce the potential for index crossover, or "tag jumping" (see Box 2; Schnell, Bohmann, & Gilbert, 2015). To address these issues, Illumina has developed a two-step PCR protocol using uniformly tailed primers across samples for the first step and sample-specific indexes for the second PCR, which could reduce bias related to index sequence variations (Berry, Mahfoudh, Wagner, & Loy, 2011; Miya et al., 2015; O'Donnell et al., 2016). Regardless of the strategy employed, extreme care is needed to ensure primer quality control (e.g., both use of small aliquots from stocks as well as proper cleaning of PCR-amplified products to remove indexing primers after amplification; Schnell, Bohmann, et al. 2015). When a species detection is suspected as highly unlikely in a sample, single-species quantitative PCR (qPCR) can be used to verify its presence from the same eDNA sample because qPCR does not suffer from the same technical sources of error. Additional suggestions for dealing with multiplexing artefacts are suggested in Box 2 under "abundance filtering."

In addition, both positive and negative controls must be used in the laboratory to ensure sample integrity (Figure 3). Use of positive control samples (either from pooled DNA extracts derived from tissue at the PCR stage, or used at the extraction stage alongside that of eDNA samples) can help evaluate sequencing efficiency and

multiplexing errors in the eDNA metabarcoding workflow (Hänfling et al., 2016; Olds et al., 2016; Port et al., 2016). Careful thought in the construction of the mock community is needed. Typically, species not expected in the study area are used (Olds et al., 2016; Thomsen et al., 2016) such that if there is contamination during the workflow, their reads can be identified, removed and serve as a control for detecting contamination when it occurs.

Negative controls should be introduced at each stage of laboratory work (i.e., filtration—if performed in the laboratory, extraction, PCR and indexing). We recommend that an equivalent amount of technical replication should be used on negative and positive controls as that carried out on actual samples (Ficetola et al., 2015). Furthermore, it is becoming important that negative controls are sequenced regardless of having detectable amounts of DNA because contamination can be below detection limits of quantification and sequences found in these controls can be used to detect demultiplexing errors or used in statistical modelling to rule out false-positive detections (Olds et al., 2016).

Finally, an important but often neglected consideration for the eDNA metabarcoding workflow is the identification of technical artefacts that arise independently of true biological variation. For example, recently in a study focused on bacterial biodiversity using the 16S locus, it was shown that a run effect can be confounded with a sample effect if it is not accounted for (e.g., by splitting sample groups across multiple Illumina runs, Chase et al., 2016); however, it remains to be seen whether such technical artefacts are also prevalent for loci used for metabarcoding plant and animals from eDNA (COI, 18S, ITS, etc.), and more research is needed. Until then, careful thought into how samples are pooled and run on a sequencer seems warranted in order to not confound the hypotheses being tested.

### 4.3 | At the keyboard

Bioinformatic processing of high-throughput sequence data sets requires the use of UNIX pipelines (or graphical wrappers of such tools, Bik et al., 2012). Metabarcoding of animal and plant community DNA is comprehensively outlined in Coissac, Riaz, and Puillandre (2012). Below and in Box 3, we highlight the common practices to community DNA metabarcoding and deviations for studies focusing on macro-organism eDNA metabarcoding.



**BOX 3 HOW TO TRANSFORM READS FROM HTS PLATFORMS INTO MEASURES OF RICHNESS****MOTU clustering**

While this step is not always necessary and depends on the target set of taxa (Lacoursière-Roussel, Dubois, Normandeau, & Bernatchez, 2016), the amplicon length sequenced (Deiner et al., 2016) and completeness of the reference database (Chain, Brown, MacIsaac, & Cristescu, 2016), clustering of sequencing reads into MOTUs is often performed prior to taxonomic assignment. MOTU clustering is the process whereby multiple reads are grouped according to set criteria of similarity based on an initial seed (Creer et al., 2016; Egan et al., 2013). Here, a centroid sequence is selected and depending on the set radius or similarity cut-off, closely related sequences are grouped under each centroid sequence (USEARCH, Edgar, 2010). The level of similarity selected depends on the study and taxon used, based on the knowledge of intraspecific diversity of the studied taxon. Commonly used cut-offs range from 97% to 99% (Bista et al., 2017; Fahner, Shokralla, Baird, & Hajibabaei, 2016; Olds et al., 2016). For example, the cut-off selected could depend on known levels of intraspecific diversity of the studied taxon, which could be estimated from an existing reference database. Some commonly used clustering algorithms include USEARCH (Edgar, 2010), VSEARCH (Rognes, Flouri, Nichols, Quince, & Mahé, 2016), CROP (Bayesian clustering algorithm; Hao, Jiang, & Chen, 2011), swarm (Mahé, Rognes, Quince, De Vargas, & Dunthorn, 2014) and mothur (an alignment-based clustering method, Schloss et al., 2009).

**Taxonomic assignment**

Identification of HTS reads is achieved through a comparison of anonymous MOTU clusters/centroid sequences or direct comparisons of reads remaining after quality filtering against a reference database. Depending on the taxon of study and the marker used, the reference database may consist of publicly available sequences or study-generated reference sequences.

The challenges of taxonomic assignment have been the subject of a considerable literature so we only briefly discuss this important aspect of the bioinformatics pipeline (e.g., Bazinet & Cummings, 2012). A number of different approaches have been suggested including assignment based on sequence similarity via alignment programs like BLAST or similarity searches using hidden Markov models such as JM-MOTU (Jones, Ghoorah, & Blaxter, 2011), MG-RAST (Glass, Wilkening, Wilke, Antonopoulos, & Meyer, 2010), sequence composition and machine learning approaches (e.g., RDP (Wang, Garrity, Tiedje, & Cole, 2007), TACOA (Diaz, Krause, Goesmann, Niehaus, & Nattkemper, 2009)), phylogenetic placement (e.g., PPLACER Matsen, Kodner, & Armbrust, 2010), probabilistic taxonomic placement (e.g., PROTAX (Somervuo, Koskela, Pennanen, Henrik Nilsson, & Ovaskainen, 2016; Somervuo et al., 2017), minimum entropy decomposition (e.g., oligotyping, Eren et al., 2015), MEGAN (Huson, Auch, Qi, & Schuster, 2007) and ecotag (Boyer et al., 2016). A number of widely used programs use combinations of these methods; for example, the program SAP (Munch, Boomsma, Huelsenbeck, Willerslev, & Nielsen, 2008) uses BLAST searches of the NCBI database and phylogenetic reconstruction to establish taxonomic identity of query sequences. Most of these methods and various derivatives are nicely discussed and compared by Bazinet and Cummings (2012). Two major determinants of the utility of these different approaches are the specific eDNA markers and the breadth and resolution of reference databases. Some markers have better representation in available databases and greater coverage of relevant species diversity. Taxonomic assignment using the BLAST algorithm (Camacho et al., 2009) is commonly used, and depending on the study, different selection criteria are specified, such as e-value, maximum ID or length of matching sequence, number of top hits selected. Caution is warranted in strictly relying on this approach, as errors in the curation of sequences in publicly available databases can propagate through the analysis and lead to misidentification of sequences. Ideally, a combination of approaches is used, and when feasible, the resultant species assignments should be vetted with independent data based on the known distribution and ecology of the species.

**Diversity analysis**

The goal of most eDNA metabarcoding studies is to accurately characterize the species richness of the community under study. Calculation of diversity indices using appropriate software allows modelling and ecological association of sequencing results. Important considerations when attempting ecological associations include appropriate data standardization to account for variations in sequencing depth and the careful selection of diversity indexes. The most common assessments include alpha diversity (rarefaction, visualization of taxonomic profiles) and beta diversity (principal components/coordinates analysis, NDMS ordination, etc.), prior to hypothesis testing via downstream statistical analysis.

Bioinformatic pipelines and parameters must be carefully considered (Box 2), and it is important to work with a knowledgeable computational researcher to understand how processing can impact the biological results and conclusions. Before computationally processing

an eDNA metabarcoding data set, perhaps the strongest message from Coissac et al. (2012) is to identify the differences between the analysis of data derived from microbial and macro-organismal groups. As microbial ecologists have been inspired to use sequence-based

identification of taxa over the past 40 years (Creer et al., 2016), the range of software solutions to analyse microbial metabarcoding data sets is unsurprisingly extensive (Bik et al., 2012). Perhaps more importantly, a number of established and maintained databases exist featuring many of the commonly used microbial taxonomic markers for prokaryotes (Cole et al., 2009), microbial eukaryotes (Guillou et al., 2013; Pruesse et al., 2007; Quast et al., 2012) and fungi (Abarenkov et al., 2010), meaning that microbial data sets can be analysed and taxonomic affiliations established are established in a straightforward way.

For macro-organism communities, preprocessing and initial quality control of eDNA metabarcoded data sets is not different from that of microbial data sets and can be acquired using packages developed either for microbial (Caporaso et al., 2010) or for macro-organism data (Boyer et al., 2016), but taxonomic assignment will require a robust data set of locus-specific reference sequences and the associated taxonomic data from a reference database (Coissac et al., 2012; Box 3). Currently, the two most common reference sources for macro-organisms are NCBI's nucleotide database (Benson et al., 2013) and the Barcode of Life Database (Ratnasingham & Hebert, 2007). The utility and taxonomic breadth of these databases can be enhanced by the creation of custom-made or hybrid databases, with the obvious additional workload and cost depending on the number of focal taxa missing from current data sources. Recently, Machida, Leray, Ho, and Knowlton (2017) have assembled and proposed metazoan mitochondrial gene sequence data sets that can be used for taxonomic assignment for environmental samples. While these data sets do not account for future growth, their methods could be repeated at the time of any new study to generate a custom reference data set for taxonomic assignment.

Macro-organism eDNA metabarcoding data sets are associated with advantages compared to microbial data sets because the number of taxa in any survey will be comparatively low, reducing the computational time needed for taxonomic annotation. Moreover, the species delimitation concepts and taxonomic markers associated with macro-organisms are well developed (de Queiroz, 2005) and can even be used to analyse population genetic structure (Sigsgaard et al., 2016; Thomsen & Willerslev, 2015), or delimit species boundaries (Coissac et al., 2012; Hebert et al., 2003; Tang, Humphreys, Fontaneto, & Barracough, 2014). Reliance on the vast knowledge we have for animal and plant taxonomy and biogeography is a distinct advantage for eDNA metabarcoding because of the independent test that it provides to calibrate and test the tool for its precision and accuracy (Deiner et al., 2016).

#### 4.4 | Data archiving for transparency

As eDNA applications continue to develop, all procedures used in the field, laboratory and during bioinformatic data processing require a strong commitment to transparency on the part of researchers (Nekrutenko & Taylor, 2012). Here, we outline best practices for eDNA metabarcoding studies of macro-organisms, following on from well-established standards in the fields of microbiology and genomics

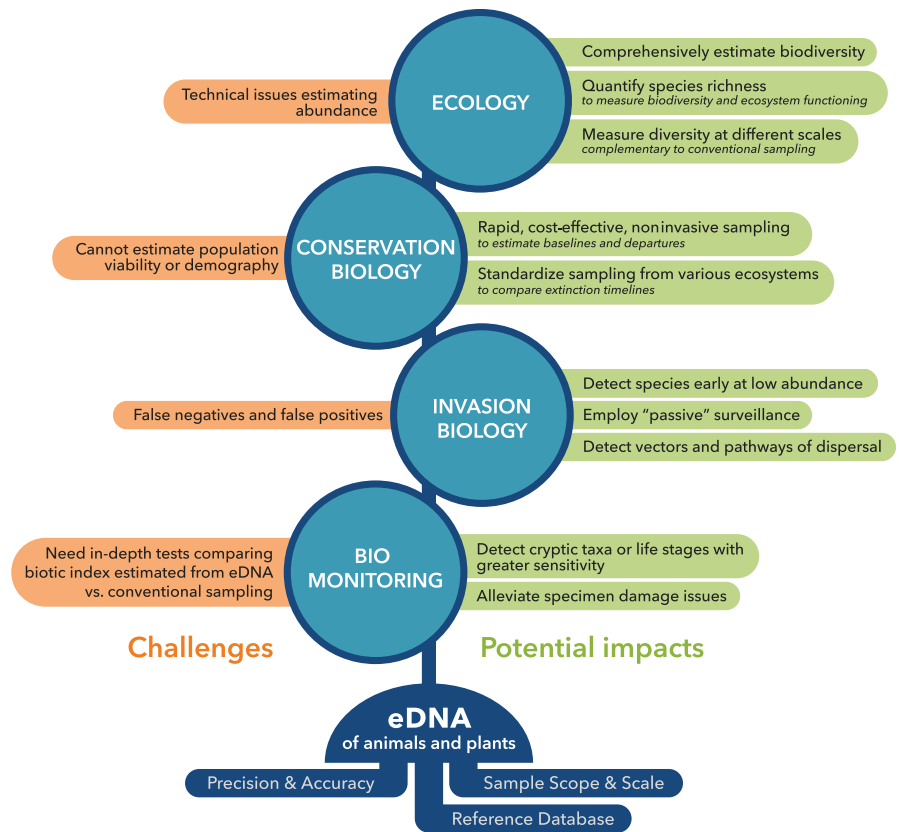
(Yilmaz et al., 2011). First, raw FASTQ files from any HTS run need to be submitted to the Sequence Read Archive (SRA) of NCBI or the European Nucleotide Archive (ENA) and other such public national databases before publication. Archiving raw data in publicly available databases is common practice in virtually all genomics and transcriptomic studies because it allows studies to be reanalysed with new computational tools and standards. In fact, archiving raw data is becoming increasingly mandatory at many evolutionary and ecology biology journals, inclusive of Molecular Ecology. Second, researchers should adhere to minimum reporting standards defined by the broader genomics community, such as the MIMARKS (minimum information about a marker gene sequence) and MIxS (minimum information about any "x" sequence) specifications (Yilmaz et al., 2011). Goldberg et al. (2016) have made specific recommendations for upholding these reporting standards specific to eDNA studies (see Table 1 in Goldberg et al., 2016).

Third, computational processing of data needs to be reproducible (Sandve, Nekrutenko, Taylor, & Hovig, 2013). For eDNA metabarcoding studies, it is increasingly common to deposit a comprehensive sample mapping file (e.g., formatted in the QIIME tab-delimited style, containing the indexes used for creating libraries so that raw data can be demultiplexed and properly trimmed) along with MOTU clustering or taxonomic binning of results, and documentation of all bioinformatics commands, in a complementary repository such as DRYAD (<http://datadryad.org/>), GITHUB (<https://github.com/github>) or FIGSHARE (<http://figshare.com>). Sandve et al. (2013) provide 10 rules that can be followed to ensure such reproducibility, and we strongly encourage researchers using eDNA metabarcoding methods to uphold these practices and take advantage of archiving intermediate steps (Box 2) of their analysis for full transparency.

## 5 | EMERGING APPLICATIONS FOR EDNA METABARCODING

### 5.1 | Applications in ecology

Quantifying the richness and abundance of species in natural communities is and will continue to be a goal in many ecological studies. Information about species richness garnered from eDNA is not necessarily different from conventional approaches (Table 1), but the scale, speed and comprehensiveness of that information is (Figure 4). For example, Drummond et al. (2015) demonstrated the near-complete analysis of biodiversity (e.g., from bacteria to animals and plants) from top soil is possible. Collection of data on this taxonomic scale opens up new opportunities with respect to measuring community composition and turnover across space and time. In addition to estimating species richness, a major area of research in ecology is determining whether observed community changes surpass acceptable thresholds for certain desired ecosystem functions (Jackson et al., 2016). Biodiversity and ecosystem functioning research requires tracking species in multiple taxonomic groups and trophic levels, along with changes in ecosystem function. Environmental DNA metabarcoding has the potential to facilitate biodiversity and ecosystem function research by



**FIGURE 4** Opportunities and challenges of using environmental DNA as a tool for assessing community structure in different fields of study. The tool is reliant on a foundation (blue half circle) of continued research to improve technological aspects and continued development of DNA-based reference libraries for the identification of sequences found in the environment

improving our knowledge of predator/prey relationships, mutualisms such as plant–pollinator interactions and food webs in highly diverse systems composed of small cryptic species (e.g., De Vere et al., 2017; Hawkins et al., 2015; Xu et al., 2015). Knowledge of species co-occurrences and interactions in these instances will additionally foster the study of metaecosystems and provide data to guide management decisions at the ecosystem scale (Bohan et al., 2017). What will remain challenging is moving beyond richness estimates to also obtaining species abundance data (Figures 2 and 4).

## 5.2 | Applications in conservation biology

Given the rapid rate at which biodiversity is declining worldwide (Butchart et al., 2010), it is critical that we improve the effectiveness of strategies to halt or reverse this loss (Thomsen & Willerslev, 2015; Valentini et al., 2016). Accordingly, developing tools that enable rapid, cost-effective and noninvasive biodiversity assessment such as eDNA metabarcoding, especially for rare and cryptic species, is paramount (Figure 4). Improved estimates of the distribution of vulnerable species, and done so noninvasively, would facilitate policy development and allow for efficient targeting of management efforts across habitats (Kelly et al., 2014; Thomsen & Willerslev, 2015). For example, documenting the presence of threatened species in a habitat can trigger a suite of actions under laws pertaining to biodiversity conservation (e.g., US Endangered Species Act). Frequently, data relevant to policy are derived from monitoring efforts mandated by environmental laws imparting a significant consequence to the data collected (Kelly et al., 2014).

Environmental DNA-based monitoring is likely to be a tremendous boon to often underfunded public agencies charged with compliance to data-demanding laws. Specifically, eDNA metabarcoding will be useful for monitoring communities when many species are of conservation concern. Vernal pools throughout California are a prime example because they contain 20 US federally listed endangered or threatened species of plants and animals. Monitoring species richness with soil and water samples from a habitat such as this would provide a comprehensive sampling method to ascertain needed community data for their conservation and management (Deiner, Hull, & May, 2017). However, while eDNA metabarcoding may be important for noninvasively gaining access to the distribution of vulnerable species, it cannot be used to differentiate between alive and dead organisms or estimate many demographic parameters important of population viability analysis (Beissinger & McCullough, 2002).

Quantifying baselines of animal and plant species richness and departures from those baselines is central to the assessment of environmental impact and conservation (Taylor & Gemmell, 2016). The application of eDNA metabarcoding methods to different samples types, which taken together allow inference across time (e.g., surface water and sediment layers from a core in a lake, Figure 1), provides a unique tool to document local extinctions and long-term changes in ecosystems. Extinction models often rely on and understanding extinction timelines (reviewed in Thomsen & Willerslev, 2015). The efficiency of eDNA metabarcoding to track the timing of extinctions associated with previous glacial events has been demonstrated in mammals (Haile et al., 2009) and plants (Willerslev et al., 2014). Thus, environmental DNA metabarcoding of different sample types

from the same site offers an excellent opportunity to better understand the extinction consequences of perturbations and could inform scenario modelling under climate change.

### 5.3 | Applications in invasion biology

Because one of the first applications of eDNA to macro-organisms was the detection of North American bullfrogs in French ponds (Ficetola et al., 2008), the method immediately came to the attention of researchers interested in invasion biology (e.g., Egan et al., 2013; Goldberg, Sepulveda, Ray, Baumgardt, & Waits, 2013; Jerde, Mahon, Chadderton, & Lodge, 2011; Takahara, Minamoto, & Doi, 2013; Tréguier et al., 2014). These initial studies, as well as much ongoing research, continue to be based on species-specific primers, where positive amplification provides occurrence evidence for a particular invasive species. In invasion biology with eDNA, such a targeted approach is referred to as “active” surveillance (Simmons, Tucker, Chadderton, Jerde, & Mahon, 2015).

On the contrary, eDNA metabarcoding makes it possible to detect the presence of many species simultaneously, including species not previously suspected of being present. This broader untargeted approach is called “passive” surveillance in management applications (Figure 4; Simmons et al., 2015). On the downside, due to a trade-off in primer specificity, we expect that eDNA metabarcoding may be less sensitive in detecting some species or that the detection rate of a species can change depending on species richness. Adopting a dual approach of passive and active surveillance could be considered in cases where the risk of a new invasion is high and where cost-effective eradication plans for undesirable species are likely to be successful (Lodge et al., 2016).

Avoiding future introductions and reducing the spread of exotic species is paramount in natural resource policy (Lodge et al., 2016). Environmental DNA metabarcoding relevant to management includes early detection of incipient invasive populations in the environment, surveillance of invasion pathways, for example, ballast water of ships (Egan et al., 2015; Zaiko et al., 2015) and the live bait trade (Mahon, Nathan, & Jerde, 2014). While eDNA metabarcoding is not yet routinely used for biosecurity regulation of invasive species or enforcement in many settings, it has the potential to become valuable monitoring tool for biological invasions. An important challenge for the use of eDNA metabarcoding in invasive species detections are false positives and false negatives as both outcomes can trigger action or inaction when not required, causing a potentially large burden on entities responsible for invasive species mitigation and control (Fig. 4). Therefore, continued research to reduce or understand the nature of false positives and false negatives will reduce uncertainty in the tool and facilitate greater adoption.

### 5.4 | Applications in biomonitoring

Pollution of air, water and land resources generated from processes such as urbanization, food production and mining is one of the many emerging global challenges we are facing in the 21st century

(Vörösmarty et al., 2010). Determining the origin, transport and effects of most pollution is challenging because it accumulates through both point sources (e.g., wastewater effluent) and diffused sources related to land-use types (e.g., agriculture or urbanization). In this context, the presence of tolerant or absence of sensitive organisms has been used to determine the consequences of pollution on ecosystem health throughout the world and is termed “biological monitoring” or “biomonitoring” (Bonada, Prat, Resh, & Statzner, 2006). The extent to which animals and plants have been used in biomonitoring depends on the unique characteristics of the taxonomic group monitored and their relationship to the pollution of interest (Bonada et al., 2006; Stankovic, Kalaba, & Stankovic, 2014). Most biomonitoring programmes take community composition and often abundance of taxa into account and calculate what is known as a biotic index (Friberg et al., 2011). Biotic indices take many forms and are typically surrogates for the impacts of pollution (e.g., SPEAR index for toxicant exposure in water, Liess, Schäfer, & Schriever, 2008).

Applying eDNA metabarcoding in the context of biomonitoring is a major avenue of research. Metabarcoding of community DNA samples has shown greater sensitivity for detecting cryptic taxa or life stages and can alleviate the problem of identifying damaged specimens of which render morphological tools ineffective (Gibson et al., 2014; Hajibabaei, Shokralla, Zhou, Singer, & Baird, 2011). These two issues alone are known to create large variances in biotic index estimation (Pfreder et al., 2010). Application of eDNA metabarcoding to animals and plants used in biomonitoring requires in-depth testing of conventional survey methods and eDNA-based approaches (Figure 4), to understand whether species richness estimates derived from the two methods result in a similar measure for the biotic index of interest or whether new biotic indices need to be development that can simultaneously consider both forms of information. Promising steps forward are being made through the DNAqua-Net COST Action (<http://dnaqua.net/>), which is a consortium of over 45 countries working together to develop genetic tools for bioassessment of aquatic ecosystems in Europe (Leese et al., 2016).

### 5.5 | Applications in citizen science and biodiversity education

The simplicity of the protocol used to collect environmental samples has created an avenue for citizen scientist programmes to be built around surveying for biodiversity using eDNA (Biggs et al., 2015). With the development of sample kits from commercial companies specifically used for eDNA analysis (e.g., GENIDAQS, ID-GENE, Jonah Ventures, NatureMetrics, Spygen), there now exists a novel opportunity to engage the public in biodiversity science, which could accompany already-established biodiversity events, such as BioBlitz (National Geographic Society). Use of eDNA metabarcoding in this context will likely provide an unprecedented tool for education and outreach about biodiversity, and increase awareness about its decline. Challenges that hinder integration of eDNA metabarcoding in citizen science projects and educational opportunities are the

time and costs needed to process samples and user-friendly data visualization tools to allow exploration of the data once provided. Thus, finding ways to cut costs and speed up data generation (a goal common for any application of the tool), as well as creation of applications for exploration of data on smart phones and desktops alike, is needed to propel the use of eDNA applications in citizen science and education.

## 6 | CONCLUSIONS

As the tool of eDNA metabarcoding continues to develop, our understanding regarding the analysis of eDNA from macro-organismal communities, including optimal field, laboratory and bioinformatics workflows, will continue to improve in the foreseeable future. Concurrently, we need to gain a better understanding of the spatial and temporal relationship between eDNA and living communities to improve precision, accuracy, and to enhance the ecological and policy relevance of eDNA (Barnes & Turner, 2016; Kelly et al., 2014). Ultimately, the errors and uncertainties associated with eDNA metabarcoding studies can often be mitigated by thoughtful study design, appropriate primer choice and robust sampling and replication: as Murray, Coghlan, and Bunce (2015) emphasize, “no amount of high-end bioinformatics can compensate for poorly prepared samples, artefacts or contamination.”

Over time, a loop in which improved eDNA metabarcoding methods reduce uncertainty about the meaning of both positive and negative eDNA detections for a species will in turn generate the motivation for continued improvements and use of eDNA metabarcoding methods, thus resulting in the adoption of eDNA metabarcoding as a comparable method for estimating species richness. We predict that over the next decade, eDNA metabarcoding of animals and plants will become a standard surveying tool that will complement conventional methods and accelerate our understanding of biodiversity across the planet.

## ACKNOWLEDGEMENTS

We thank Nigel G. Yoccoz Tromsø and three additional anonymous reviewers whose feedback was valuable in revising our manuscript. We thank Kristina Davis for help in drafting figures. Manuscript collaboration was facilitated by the National Science Foundation through the Coastal SEES Grant to DML (EF-1427157; KD, DML, MEP), a DoD SERDP Grant to DML (W912HQ-12-C-0073 (RC-2240); KD, DML, MEP) and a NOAA CSCOR Grant to DML (KD, DML). National Science Foundation Research Coordination Network award to HMB (DBI-1262480) supported a related eDNA-focused Symposium at the 2016 Annual Meeting of the Ecological Society of America organized by KD, DML and MEP. In addition, this article is based upon work from COST Action DNAqua-Net (CA15219; KD, FA, SC), supported by the COST (European Cooperation in Science and Technology) programme, by the Swiss National Science Foundation Grant No PP00P3\_150698 (to FA) and Eawag (FA and EM); the

Natural Environment Research Council (NERC): NBAF Pilot Project Grant (NBAF824 2013-14), Standard Grant PollerGEN (NE/N003756/1) and Highlight Topic Grant LOFRESH (NE/N006216/1) and the Freshwater Biological Association (FBA; Gilson Le Cren Memorial Award 2014). IB was funded by a Knowledge Economy Skills Scholarship (KESS), a pan-Wales higher-level skills initiative led by Bangor University on behalf of the HE sector in Wales. It is part funded by the Welsh Government's European Social Fund (ESF) convergence programme for West Wales and the Valleys.

## DATA ACCESSIBILITY

No data are associated with the manuscript.

## AUTHOR CONTRIBUTIONS

K.D. outlined and edited the review. All authors contributed at least one section of primary writing and contributed to editing of the manuscript. K.D., H.M.B and E.M. synthesized sections and drafted figures.

## ORCID

Kristy Deiner  <http://orcid.org/0000-0001-9755-3223>

Holly M. Bik  <http://orcid.org/0000-0002-4356-3837>

Anaïs Lacoursière-Roussel  <http://orcid.org/0000-0002-9345-5682>

Florian Altermatt  <http://orcid.org/0000-0002-4831-6958>

Simon Creer  <http://orcid.org/0000-0003-3124-3550>

Natasha de Vere  <http://orcid.org/0000-0001-9593-6925>

## REFERENCES

- Abarenkov, K., Henrik Nilsson, R., Larsson, K. H., Alexander, I. J., Eberhardt, U., Erland, S., ... Pennanen, T. (2010). The UNITE database for molecular identification of fungi—recent updates and future perspectives. *New Phytologist*, 186, 281–285.
- Andersen, K., Bird, K. L., Rasmussen, M., Haile, J., Breuning-Madsen, H., Kjaer, K. H., ... Willerslev, E. (2012). Meta-barcoding of 'dirt' DNA from soil reflects vertebrate biodiversity. *Molecular Ecology*, 21, 1966–1979.
- Anderson-Carpenter, L. L., McLachlan, J. S., Jackson, S. T., Kuch, M., Lumibao, C. Y., & Poinar, H. N. (2011). Ancient DNA from lake sediments: Bridging the gap between paleoecology and genetics. *BMC Evolutionary Biology*, 11, 30.
- Andruszkiewicz, E. A., Starks, H. A., Chavez, F. P., Sassoubre, L. M., Block, B. A., & Boehm, A. B. (2017). Biomonitoring of marine vertebrates in Monterey Bay using eDNA metabarcoding. *PLoS ONE*, 12, e0176343.
- Barnes, M. A., & Turner, C. R. (2016). The ecology of environmental DNA and implications for conservation genetics. *Conservation Genetics*, 17, 1–17.
- Bazin, A. L., & Cummings, M. P. (2012). A comparative evaluation of sequence classification programs. *BMC Bioinformatics*, 13, 1.
- Beissinger, S. R., & McCullough, D. R. (2002). *Population viability analysis*. Chicago: The University of Chicago Press.
- Bell, K. L., de Vere, N., Keller, A., Richardson, R. T., Gous, A., Burgess, K. S., & Brosi, B. J. (2016). Pollen DNA barcoding: Current applications and future prospects. *Genome*, 59, 629–640.



- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2013). GenBank. *Nucleic Acids Research*, 41, D36–D42.
- Berry, D., Mahfoudh, K. B., Wagner, M., & Loy, A. (2011). Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. *Applied and Environmental Microbiology*, 77, 7846–7849.
- Bienert, F., De Danieli, S., Miquel, C., Coissac, E., Poillot, C., BRUN, J.J., & Taberlet, P. (2012). Tracking earthworm communities from soil DNA. *Molecular Ecology*, 21, 2017–2030.
- Biggs, J., Ewald, N., Valentini, A., Gaboriaud, C., Dejean, T., Griffiths, R. A., ... Brotherton, P. (2015). Using eDNA to develop a national citizen science-based monitoring programme for the great crested newt (*Triturus cristatus*). *Biological Conservation*, 183, 19–28.
- Bik, H. M., Porazinska, D. L., Creer, S., Caporaso, J. G., Knight, R., & Thomas, W. K. (2012). Sequencing our way towards understanding global eukaryotic biodiversity. *Trends in Ecology & Evolution*, 27, 233–243.
- Bista, I., Carvalho, G., Walsh, K., Seymour, M., Hajibabaei, M., Lallias, D., ... Creer, S. (2017). Annual time-series analysis of aqueous eDNA reveals ecologically relevant dynamics of lake ecosystem biodiversity. *Nature Communications*, 8, 14087.
- Blaxter, M., Mann, J., Chapman, T., Thomas, F., Whitton, C., Floyd, R., & Abebe, E. (2005). Defining operational taxonomic units using DNA barcode data. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 360, 1935–1943.
- Bohan, D. A., Vacher, C., Tamaddon-Nezhad, A., Raybould, A., Dumbrell, A. J., & Woodward, G. (2017). Next-generation global biomonitoring: Large-scale, automated reconstruction of ecological networks. *Trends in Ecology & Evolution*, 32, 477–487.
- Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., ... Caporaso, J. G. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods*, 10, 57–59.
- Bonada, N., Prat, N., Resh, V. H., & Statzner, B. (2006). Developments in aquatic insect biomonitoring: A comparative analysis of recent approaches. *Annual Review of Entomology*, 51, 495–523.
- Bortolus, A. (2008). Error cascades in the biological sciences: The unwanted consequences of using bad taxonomy in ecology. *AMBIO: A Journal of the Human Environment*, 37, 114–118.
- Boyer, S., Cruickshank, R. H., & Wratten, S. D. (2015). Faeces of generalist predators as 'biodiversity capsules': A new tool for biodiversity assessment in remote and inaccessible habitats. *Food Webs*, 3, 1–6.
- Boyer, F., Mercier, C., Bonin, A., Le Bras, Y., Taberlet, P., & Coissac, E. (2016). obitools: A unix-inspired software package for DNA metabarcoding. *Molecular Ecology Resources*, 16, 176–182.
- Brown, S.K., & Blois, J.L. (2016). Ecological Insights from Ancient DNA, In *eLS* (pp. 1–7). Chichester: John Wiley & Sons, Ltd.
- Butchart, S. H., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J. P., Almond, R. E., ... Bruno, J. (2010). Global biodiversity: Indicators of recent declines. *Science*, 328, 1164–1168.
- Calvignac-Spencer, S., Merkel, K., Kutzner, N., Kühl, H., Boesch, C., Kappler, P. M., ... Leendertz, F. H. (2013). Carrion fly-derived DNA as a tool for comprehensive and cost-effective assessment of mammalian biodiversity. *Molecular Ecology*, 22, 915–924.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and applications. *BMC Bioinformatics*, 10, 1.
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., ... Gordon, J. I. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335–336.
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-lyons, D., Lozupone, C. A., Turnbaugh, P. J., ... Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 4516–4522.
- Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., ... Wardle, D. A. (2012). Biodiversity loss and its impact on humanity. *Nature*, 486, 59–67.
- Chain, F. J. J., Brown, E. A., MacIsaac, H. J., & Cristescu, M. E. (2016). Metabarcoding reveals strong spatial structure and temporal turnover of zooplankton communities among marine and freshwater ports. *Diversity and Distributions*, 22, 493–504.
- Chase, J., Fouquier, J., Zare, M., Sonderegger, D.L., Knight, R., Kelley, S.T., ... Caporaso, J.G. (2016). Geography and location are the primary drivers of office microbiome composition. *mSystems*, 1, e00022–00016.
- Chen, S., Yao, H., Han, J., Liu, C., Song, J., Shi, L., ... Pang, X. (2010). Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE*, 5, e8613.
- Civade, R., Dejean, T., Valentini, A., Roset, N., Raymond, J.-C., Bonin, A., ... Pont, D. (2016). Spatial representativeness of environmental DNA metabarcoding signal for fish biodiversity assessment in a natural freshwater system. *PLoS ONE*, 11, e0157366.
- Coissac, E., Riaz, T., & Puillandre, N. (2012). Bioinformatic challenges for DNA metabarcoding of plants and animals. *Molecular Ecology*, 21, 1834–1847.
- Cole, J. R., Wang, Q., Cardenas, E., Fish, J., Chai, B., Farris, R. J., ... Garrity, G. M. (2009). The Ribosomal Database Project: Improved alignments and new tools for rRNA analysis. *Nucleic Acids Research*, 37, D141–D145.
- Craine, J. M., Barberán, A., Lynch, R. C., Menninger, H. L., Dunn, R. R., & Fierer, N. (2017). Molecular analysis of environmental plant DNA in house dust across the United States. *Aerobiologia*, 33, 71–86.
- Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W. K., ... Bik, H. M. (2016). The ecologist's field guide to sequence-based identification of biodiversity. *Methods in Ecology and Evolution*, 7, 1008–1018.
- Danielopol, D. L., Pospisil, P., & Rouch, R. (2000). Biodiversity in groundwater: A large-scale view. *Trends in Ecology & Evolution*, 15, 223–224.
- De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., & Taberlet, P. (2014). DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: Application to omnivorous diet. *Molecular Ecology Resources*, 14, 306–323.
- De Queiroz, K. (2005). Different species problems and their resolution. *BioEssays*, 27, 1263–1269.
- De Vere, N., Jones, L. E., Gilmore, T., Moscrop, J., Lowe, A., Smith, D., ... Ford, C. R. (2017). Using DNA metabarcoding to investigate honey bee foraging reveals limited flower use despite high floral availability. *Scientific Reports*, 7, 42838.
- Deagle, B. E., Eveson, J. P., & Jarman, S. N. (2006). Quantification of damage in DNA recovered from highly degraded samples—a case study on DNA in faeces. *Frontiers in Zoology*, 3, 11.
- Deagle, B. E., Jarman, S. N., Coissac, E., Pompanon, F., & Taberlet, P. (2014). DNA metabarcoding and the cytochrome c oxidase subunit I marker: Not a perfect match. *Biology Letters*, 10, 20140562.
- Deiner, K., & Altermatt, F. (2014). Transport distance of invertebrate environmental DNA in a natural river. *PLoS ONE*, 9, e88786.
- Deiner, K., Fronhofer, E.A., Mächler, E., Walser, J.-C., & Altermatt, F. (2016). Environmental DNA reveals that rivers are conveyor belts of biodiversity information. *Nature Communications*, 7, 12544.
- Deiner, K., Hull, J. M., & May, B. (2017). Range-wide phylogeographic structure of the vernal pool fairy shrimp (*Branchinecta lynchi*). *PLoS ONE*, 12, e0176266.
- Deiner, K., Renshaw, M. A., Li, Y., Olds, B. P., Lodge, D. M., & Pfrender, M. E. (2017). Long-range PCR allows sequencing of mitochondrial genomes from environmental DNA. *Methods in Ecology and Evolution*, 00, 1–11. <https://doi.org/10.1111/2041-210X.12836>
- Deiner, K., Walser, J.-C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, 183, 53–63.



- Diaz, N. N., Krause, L., Goesmann, A., Niehaus, K., & Nattkemper, T. W. (2009). TACOA—Taxonomic classification of environmental genomic fragments using a kernelized nearest neighbor approach. *BMC Bioinformatics*, 10, 56.
- Doi, H., Inui, R., Akamatsu, Y., Kanno, K., Yamanaka, H., Takahara, T., & Minamoto, T. (2017). Environmental DNA analysis for estimating the abundance and biomass of stream fish. *Freshwater Biology*, 62, 30–39.
- Dowle, E. J., Pochon, X., C. Banks, J., Shearer, K., & Wood, S. A. (2016). Targeted gene enrichment and high-throughput sequencing for environmental biomonitoring: A case study using freshwater macroinvertebrates. *Molecular Ecology Resources*, 16, 1240–1254.
- Drummond, A. J., Newcomb, R. D., Buckley, T. R., Xie, D., Dopheide, A., Potter, B. C., ... Grosser, S. (2015). Evaluating a multigene environmental DNA approach for biodiversity assessment. *GigaScience*, 4, 1.
- Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460–2461.
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27, 2194–2200.
- Egan, S. P., Barnes, M. A., Hwang, C. T., Mahon, A. R., Feder, J. L., Ruggiero, S. T., ... Lodge, D. M. (2013). Rapid invasive species detection by combining environmental DNA with light transmission spectroscopy. *Conservation Letters*, 6, 402–409.
- Egan, S. P., Grey, E., Olds, B., Feder, J. L., Ruggiero, S. T., Tanner, C. E., & Lodge, D. M. (2015). Rapid molecular detection of invasive species in ballast and harbor water by integrating environmental DNA and light transmission spectroscopy. *Environmental Science & Technology*, 49, 4113–4121.
- Elbrecht, V., & Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass—sequence relationships with an innovative metabarcoding protocol. *PLoS ONE*, 10, e0130324.
- Elbrecht, V., & Leese, F. (2017). PrimerMiner: An R package for development and in silico validation of DNA metabarcoding primers. *Methods in Ecology and Evolution*, 8, 622–626.
- Erb, L. A., Willey, L. L., Johnson, L. M., Hines, J. E., & Cook, R. P. (2015). Detecting long-term population trends for an elusive reptile species. *The Journal of Wildlife Management*, 79, 1062–1071.
- Eren, A. M., Morrison, H. G., Lescault, P. J., Reveillaud, J., Vineis, J. H., & Sogin, M. L. (2015). Minimum entropy decomposition: Unsupervised oligotyping for sensitive partitioning of high-throughput marker gene sequences. *The ISME Journal*, 9, 968–979.
- Evans, N. T., Li, Y., Renshaw, M. A., Olds, B. P., Deiner, K., Turner, C. R., ... Pfrender, M. E. (2017). Fish community assessment with eDNA metabarcoding: Effects of sampling design and bioinformatic filtering. *Canadian Journal of Fisheries and Aquatic Sciences*, 74, 1362–1374.
- Evans, N. T., Olds, B. P., Renshaw, M. A., Turner, C. R., Li, Y., Jerde, C. L., ... Lodge, D. M. (2016). Quantification of mesocosm fish and amphibian species diversity via environmental DNA metabarcoding. *Molecular Ecology Resources*, 16, 29–41.
- Fahner, N. A., Shokralla, S., Baird, D. J., & Hajibabaei, M. (2016). Large-Scale Monitoring of Plants through Environmental DNA Metabarcoding of Soil: Recovery, Resolution, and Annotation of Four DNA Markers. *PLoS ONE*, 11, e0157505.
- Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, 4, 423–425.
- Ficetola, G. F., Pansu, J., Bonin, A., Coissac, E., Giguët-Covex, C., De Barba, M., ... Taberlet, P. (2015). Replication levels, false presences and the estimation of the presence/absence from eDNA metabarcoding data. *Molecular Ecology Resources*, 15, 543–556.
- Foot, A. D., Thomsen, P. F., Sveegaard, S., Wahlberg, M., Kielgast, J., Kyhn, L. A., ... Gilbert, M. T. P. (2012). Investigating the potential use of environmental DNA (eDNA) for genetic monitoring of marine mammals. *PLoS ONE*, 7, e41781.
- Freeland, J. R. (2016). The importance of molecular markers and primer design when characterizing biodiversity from environmental DNA. *Genome*, 60, 358–374.
- Friberg, N., Bonada, N., Bradley, D. C., Dunbar, M. J., Edwards, F. K., Grey, J., ... Trimmer, M. (2011). Biomonitoring of human impacts in freshwater ecosystems: The good, the bad and the ugly. In G. Woodward (Ed.), *Advances in ecological research* (pp. 1–68). London: Academic Press.
- Furlan, E. M., Gleeson, D., Hardy, C. M., & Duncan, R. P. (2016). A framework for estimating the sensitivity of eDNA surveys. *Molecular Ecology Resources*, 16, 641–654.
- Gardham, S., Hose, G. C., Stephenson, S., & Chariton, A. A. (2014). DNA metabarcoding meets experimental ecotoxicology: Advancing knowledge on the ecological effects of copper in freshwater ecosystems. *Advances in Ecological Research*, 51, 79–104.
- Gibson, J. F., Shokralla, S., Curry, C., Baird, D. J., Monk, W. A., King, I., & Hajibabaei, M. (2015). Large-scale biomonitoring of remote and threatened ecosystems via high-throughput sequencing. *PLoS ONE*, 10, e0138432.
- Gibson, J., Shokralla, S., Porter, T. M., King, I., van Konynenburg, S., Janzen, D. H., ... Hajibabaei, M. (2014). Simultaneous assessment of the macrobiome and microbiome in a bulk sample of tropical arthropods through DNA metasytematics. *Proceedings of the National Academy of Sciences of the United States of America*, 111, 8007–8012.
- Giersch, J. J., Hotaling, S., Kovach, R. P., Jones, L. A., & Muhlfeld, C. C. (2017). Climate-induced glacier and snow loss imperils alpine stream insects. *Global Change Biology*, 23, 2577–2589.
- Giguët-Covex, C., Pansu, J., Arnaud, F., Rey, P.-J., Griggo, C., Gelly, L., ... Choler, P. (2014). Long livestock farming history and human landscape shaping revealed by lake sediment DNA. *Nature Communications*, 5, 3211.
- Glass, E. M., Wilkening, J., Wilke, A., Antonopoulos, D., & Meyer, F. (2010). Using the metagenomics RAST server (MG-RAST) for analyzing shotgun metagenomes. *Cold Spring Harbor Protocols*, 2010, pdb.prot5368.
- Gleick, P. (1993). *Water in crisis: A guide to the world's fresh water resources*. New York, New York: Oxford University Press.
- Goldberg, C. S., Sepulveda, A., Ray, A., Baumgardt, J., & Waits, L. P. (2013). Environmental DNA as a new method for early detection of New Zealand mudsnails (*Potamopyrgus antipodarus*). *Freshwater Science*, 32, 792–800.
- Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., ... Cornman, R. S. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, 7, 1299–1307.
- Gorički, Š., Stanković, D., Snoj, A., Kuntner, M., Jeffery, W. R., Trontelj, P., ... Aljančić, G. (2017). Environmental DNA in subterranean biology: Range extension and taxonomic implications for *Proteus*. *Scientific Reports*, 7, 45054.
- Gotelli, N. J., & Colwell, R. K. (2011). Estimating species richness. *Biological Diversity: Frontiers in Measurement and Assessment*, 12, 39–54.
- Guardiola, M., Wangenstein, O. S., Taberlet, P., Coissac, E., Uriz, M. J., & Turon, X. (2016). Spatio-temporal monitoring of deep-sea communities using metabarcoding of sediment DNA and RNA. *PeerJ*, 4, e2807.
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., ... Christen, R. (2013). The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Research*, 41, D597–D604.
- Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Gianoukos, G., ... Sodergren, E. (2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Research*, 21, 494–504.

- Haile, J., Froese, D. G., MacPhee, R. D., Roberts, R. G., Arnold, L. J., Reyes, A. V., ... Robinson, S. (2009). Ancient DNA reveals late survival of mammoth and horse in interior Alaska. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 22352–22357.
- Hajibabaei, M., Shokralla, S., Zhou, X., Singer, G. A., & Baird, D. J. (2011). Environmental barcoding: A next-generation sequencing approach for biomonitoring applications using river benthos. *PLoS ONE*, 6, e17497.
- Hänfling, B., Lawson Handley, L., Read, D. S., Hahn, C., Li, J., Nichols, P., ... Winfield, I. J. (2016). Environmental DNA metabarcoding of lake fish communities reflects long-term data from established survey methods. *Molecular Ecology*, 25, 3101–3119.
- Hao, X., Jiang, R., & Chen, T. (2011). Clustering 16S rRNA for OTU prediction: A method of unsupervised Bayesian clustering. *Bioinformatics*, 27, 611–618.
- Haouchar, D., Haile, J., McDowell, M. C., Murray, D. C., White, N. E., Allcock, R. J., ... Bunce, M. (2014). Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocene–Holocene cave deposit on Kangaroo Island, South Australia. *Quaternary Science Reviews*, 84, 56–64.
- Hawkins, J., de Vere, N., Griffith, A., Ford, C. R., Allainguillaume, J., Hegarty, M. J., ... Adams-Groom, B. (2015). Using DNA metabarcoding to identify the floral composition of honey: A new tool for investigating honey bee foraging preferences. *PLoS ONE*, 10, e0134735.
- Hebert, P. D., Ratnasingham, S., & de Waard, J. R. (2003). Barcoding animal life: Cytochrome c oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biological Sciences*, 270, S96–S99.
- Hollingsworth, P.M., Group, C.P.W., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., ... Erickson, D.L. (2009). DNA barcode for land plants. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 12794–12797.
- Hopken, M. W., Orning, E. K., Young, J. K., & Piaggio, A. J. (2016). Molecular forensics in avian conservation: A DNA-based approach for identifying mammalian predators of ground-nesting birds and eggs. *BMC Research Notes*, 9, 1.
- Hunter, M. E., Oyler-McCance, S. J., Dorazio, R. M., Fike, J. A., Smith, B. J., Hunter, C. T., ... Hart, K. M. (2015). Environmental DNA (eDNA) Sampling Improves Occurrence and Detection Estimates of Invasive Burmese Pythons. *PLoS ONE*, 10, e0121655.
- Huson, D. H., Auch, A. F., Qi, J., & Schuster, S. C. (2007). MEGAN analysis of metagenomic data. *Genome Research*, 17, 377–386.
- Jackson, M.C., Weyl, O.L.F., Altermatt, F., Durance, I., Friberg, N., Dumbrell, A.J., ... Woodward, G. (2016). Chapter twelve - Recommendations for the next generation of global freshwater biological monitoring tools. In R.L.K. Alex, J. Dumbrell & W. Guy (Eds.), *Advances in ecological research* (pp. 615–636). New York: Academic Press.
- Jerde, C. L., Mahon, A. R., Chadderton, W. L., & Lodge, D. M. (2011). "Sight-unseen" detection of rare aquatic species using environmental DNA. *Conservation Letters*, 4, 150–157.
- Jo, T., Murakami, H., Masuda, R., Sakata, M. K., Yamamoto, S., & Yamamoto, T. (2017). Rapid degradation of longer DNA fragments enables the improved estimation of distribution and biomass using environmental DNA. *Molecular Ecology Resources*, 00, 1–9. <https://doi.org/10.1111/1755-0998.12685>
- Jones, M., Ghoorah, A., & Blaxter, M. (2011). jMOTU and Taxonator: Turning DNA barcode sequences into annotated operational taxonomic units. *PLoS ONE*, 6, e19259.
- Jørgensen, T., Kjær, K. H., Haile, J., Rasmussen, M., Boessenkool, S., Andersen, K., ... Orlando, L. (2012). Islands in the ice: Detecting past vegetation on Greenlandic nunataks using historical records and sedimentary ancient DNA Meta-barcoding. *Molecular Ecology*, 21, 1980–1988.
- Kao, C.-M., Liao, H.-Y., Chien, C.-C., Tseng, Y.-K., Tang, P., Lin, C.-E., & Chen, S.-C. (2016). The change of microbial community from chlorinated solvent-contaminated groundwater after biostimulation using the metagenome analysis. *Journal of Hazardous Materials*, 302, 144–150.
- Kelly, R. P. (2016). Making environmental DNA count. *Molecular Ecology Resources*, 16, 10–12.
- Kelly, R. P., Port, J. A., Yamahara, K. M., Martone, R. G., Lowell, N., Thomsen, P. F., ... Caldwell, M. R. (2014). Harnessing DNA to improve environmental management. *Science*, 344, 1455–1456.
- Klymus, K. E., Richter, C. A., Chapman, D. C., & Paukert, C. (2015). Quantification of eDNA shedding rates from invasive bighead carp *Hypophthalmichthys nobilis* and silver carp *Hypophthalmichthys molitrix*. *Biological Conservation*, 183, 77–84.
- Kraaijeveld, K., Weger, L. A., Ventayol García, M., Buermans, H., Frank, J., Hiemstra, P. S., & Dunnen, J. T. (2015). Efficient and sensitive identification and quantification of airborne pollen using next-generation DNA sequencing. *Molecular Ecology Resources*, 15, 8–16.
- Kunin, V., Engelbrektson, A., Ochman, H., & Hugenholtz, P. (2010). Wrinkles in the rare biosphere: Pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology*, 12, 118–123.
- Lacoursière-Roussel, A., Côté, G., Leclerc, V., & Bernatchez, L. (2016). Quantifying relative fish abundance with eDNA: A promising tool for fisheries management. *Journal of Applied Ecology*, 53, 1148–1157.
- Lacoursière-Roussel, A., Dubois, Y., Normandeau, E., & Bernatchez, L. (2016). Improving herpetological surveys in eastern North America using the environmental DNA method. *Genome*, 59, 991–1007.
- Lacoursière-Roussel, A., Rosabal, M., & Bernatchez, L. (2016). Estimating fish abundance and biomass from eDNA concentrations: Variability among capture methods and environmental conditions. *Molecular Ecology Resources*, 16, 1401–1414.
- Leese, F., Altermatt, F., Bouchez, A., Ekrem, T., Hering, D., Csabai, Z. S., ... Várbíró, G. (2016). DNAqua-Net: Developing new genetic tools for bioassessment and monitoring of aquatic ecosystems in Europe. *Research Ideas and Outcomes (RIO)*, 2, e11321.
- Lejzerowicz, F., Esling, P., Majewski, W., Szczuciński, W., Decelle, J., Obadia, C., ... Pawlowski, J. (2013). Ancient DNA complements microfossil record in deep-sea subsurface sediments. *Biology Letters*, 9, 20130283.
- Leray, M., & Knowlton, N. (2017). Random sampling causes the low reproducibility of rare eukaryotic OTUs in Illumina COI metabarcoding. *PeerJ*, 5, e3006.
- Leray, M., Yang, J. Y., Meyer, C. P., Mills, S. C., Agudelo, N., Ranwez, V., ... Machida, R. J. (2013). A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents. *Frontiers in Zoology*, 10, 34.
- Liess, M., Schäfer, R. B., & Schriever, C. A. (2008). The footprint of pesticide stress in communities—species traits reveal community effects of toxicants. *Science of the Total Environment*, 406, 484–490.
- Lim, N. K., Tay, Y. C., Srivathsan, A., Tan, J. W., Kwik, J. T., Baloglu, B., ... Yeo, D. C. (2016). Next-generation freshwater bioassessment: eDNA metabarcoding with a conserved metazoan primer reveals species-rich and reservoir-specific communities. *Royal Society Open Science*, 3, 160635.
- Lodge, D. M., Simonin, P. W., Burgiel, S. W., Keller, R. P., Bossenbroek, J. M., Jerde, C. L., ... Wittmann, M. E. (2016). Risk analysis and bioeconomics of invasive species to inform policy and management. *Annual Review of Environment and Resources*, 41, 453–488.
- Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J., & Pallen, M. J. (2012). Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology*, 30, 434–439.
- Machida, R.J., Leray, M., Ho, S.-L., & Knowlton, N. (2017). Metazoan mitochondrial gene sequence reference datasets for taxonomic assignment of environmental samples. *Scientific Data*, 4, 170027.
- Mahé, F., Rognes, T., Quince, C., De Vargas, C., & Dunthorn, M. (2014). Swarm: Robust and fast clustering method for amplicon-based studies. *PeerJ*, 2, e593.

- Mahon, A. R., Nathan, L. R., & Jerde, C. L. (2014). Meta-genomic surveillance of invasive species in the bait trade. *Conservation Genetics Resources*, 6, 563–567.
- Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., ... Chen, Z. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437, 376–380.
- Matsen, F. A., Kodner, R. B., & Armbrust, E. V. (2010). pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11, 1.
- McGee, K. M., & Eaton, W. D. (2015). A comparison of the wet and dry season DNA-based soil invertebrate community characteristics in large patches of the bromeliad *Bromelia pinguin* in a primary forest in Costa Rica. *Applied Soil Ecology*, 87, 99–107.
- Merkes, C. M., McCalla, S. G., Jensen, N. R., Gaikowski, M. P., & Amberg, J. J. (2014). Persistence of DNA in carcasses, slime and avian feces may affect interpretation of environmental DNA data. *PLoS ONE*, 9, e113346.
- Minamoto, T., Yamanaka, H., Takahara, T., Honjo, M. N., & Kawabata, Z. I. (2012). Surveillance of fish species composition using environmental DNA. *Limnology*, 13, 193–197.
- Miya, M., Sato, Y., Fukunaga, T., Sado, T., Poulsen, J., Sato, K., ... Araki, H. (2015). MiFish, a set of universal PCR primers for metabarcoding environmental DNA from fishes: Detection of more than 230 sub-tropical marine species. *Open Science*, 2, 150088.
- Moyer, G. R., Díaz-Ferguson, E., Hill, J. E., & Shea, C. (2014). Assessing Environmental DNA Detection in Controlled Lentic Systems. *PLoS ONE*, 9, e103767.
- Munch, K., Boomsma, W., Huelsenbeck, J. P., Willerslev, E., & Nielsen, R. (2008). Statistical assignment of DNA sequences using Bayesian phylogenetics. *Systematic Biology*, 57, 750–757.
- Murray, D. C., Coghlan, M. L., & Bunce, M. (2015). From benchtop to desktop: Important considerations when designing amplicon sequencing workflows. *PLoS ONE*, 10, e0124671.
- Murray, D. C., Pearson, S. G., Fullagar, R., Chase, B. M., Houston, J., Atchison, J., ... Macphail, M. (2012). High-throughput sequencing of ancient plant and mammal DNA preserved in herbivore middens. *Quaternary Science Reviews*, 58, 135–145.
- Nekrutenko, A., & Taylor, J. (2012). Next-generation sequencing data interpretation: Enhancing reproducibility and accessibility. *Nature Reviews Genetics*, 13, 667–672.
- Nichols, R. V., KOeNIGSSON, H., Danell, K., & Spong, G. (2012). Browsed twig environmental DNA: Diagnostic PCR to identify ungulate species. *Molecular Ecology Resources*, 12, 983–989.
- O'Donnell, J. L., Kelly, R. P., Lowell, N. C., & Port, J. A. (2016). Indexed PCR primers induce template-specific bias in large-scale DNA sequencing studies. *PLoS ONE*, 11, e0148698.
- O'Donnell, J. L., Kelly, R. P., Shelton, A. O., Samhouri, J. F., Lowell, N. C., & Williams, G. D. (2017). Spatial distribution of environmental DNA in a nearshore marine habitat. *PeerJ*, 5, e3044.
- Olds, B. P., Jerde, C. L., Renshaw, M. A., Li, Y., Evans, N. T., Turner, C. R., ... Shirey, P. D. (2016). Estimating species richness using environmental DNA. *Ecology and Evolution*, 6, 4214–4226.
- Pansu, J., De Danieli, S., Puissant, J., Gonzalez, J.-M., Gielly, L., Cordonnier, T., ... Taberlet, P. (2015). Landscape-scale distribution patterns of earthworms inferred from soil DNA. *Soil Biology and Biochemistry*, 83, 100–105.
- Pansu, J., Giguët-Covex, C., Ficetola, G. F., Gielly, L., Boyer, F., Zinger, L., ... Choler, P. (2015). Reconstructing long-term human impacts on plant communities: An ecological approach based on lake sediment DNA. *Molecular Ecology*, 24, 1485–1498.
- Parducci, L., Matetovici, I., Fontana, S. L., Bennett, K. D., Suyama, Y., Haile, J., ... Willerslev, E. (2013). Molecular-and pollen-based vegetation analysis in lake sediments from central Scandinavia. *Molecular Ecology*, 22, 3511–3524.
- Pawlowski, J., Esling, P., Lejzerowicz, F., Cedhagen, T., & Wilding, T. A. (2014). Environmental monitoring through protist next-generation sequencing metabarcoding: Assessing the impact of fish farming on benthic foraminifera communities. *Molecular Ecology Resources*, 14, 1129–1140.
- Pedersen, M. W., Overballe-Petersen, S., Ermini, L., Der Sarkissian, C., Haile, J., Hellstrom, M., ... Cappellini, E. (2015). Ancient and modern environmental DNA. *Philosophical Transactions of the Royal Society of London. Series B*, 370, 20130383.
- Pedersen, M. W., Ruter, A., Schweger, C., Friebe, H., Staff, R. A., Kjeldsen, K. K., ... Willerslev, E. (2016). Postglacial viability and colonization in North America's ice-free corridor. *Nature*, 537, 45–49.
- Pfrender, M., Hawkins, C., Bagley, M., Courtney, G., Creutzburg, B., Epler, J., ... Jackson, S. (2010). Assessing macroinvertebrate biodiversity in freshwater ecosystems: Advances and challenges in DNA-based approaches. *The Quarterly Review of Biology*, 85, 319–340.
- Pilliod, D. S., Goldberg, C. S., Arkle, R. S., & Waits, L. P. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, 70, 1123–1130.
- Piñol, J., Mir, G., Gomez-Polo, P., & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Molecular Ecology Resources*, 15, 819–830.
- Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., ... Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, 25, 527–541.
- Price, S. J., Eskew, E. A., Cecala, K. K., Browne, R. A., & Dorcas, M. E. (2012). Estimating survival of a streamside salamander: Importance of temporary emigration, capture response, and location. *Hydrobiologia*, 679, 205–215.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glöckner, F. O. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic acids Research*, 35, 7188–7196.
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic acids Research*, 41, D590–D596.
- Ratnasingham, S., & Hebert, P. D. (2007). BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Molecular Ecology Notes*, 7, 355–364.
- Rees, H. C., Bishop, K., Middleditch, D. J., Patmore, J. R. M., Maddison, B. C., & Gough, K. C. (2014). The application of eDNA for monitoring of the Great Crested Newt in the UK. *Ecology and Evolution*, 4, 4023–4032.
- Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R., & Gough, K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA—a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, 51, 1450–1459.
- Renshaw, M. A., Olds, B. P., Jerde, C. L., McVeigh, M. M., & Lodge, D. M. (2015). The room temperature preservation of filtered environmental DNA samples and assimilation into a phenol-chloroform-isomyl alcohol DNA extraction. *Molecular Ecology Resources*, 15, 168–176.
- Rodgers, T. W., & Janečka, J. E. (2013). Applications and techniques for non-invasive faecal genetics research in felid conservation. *European Journal of Wildlife Research*, 59, 1–16.
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A versatile open source tool for metagenomics. *PeerJ*, 4, e2584.
- Sandve, G. K., Nekrutenko, A., Taylor, J., & Hovig, E. (2013). Ten simple rules for reproducible computational research. *PLoS Computational Biology*, 9, e1003285.

- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., ... Robinson, C. J. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75, 7537–7541.
- Schmelzle, M. C., & Kinziger, A. P. (2016). Using occupancy modelling to compare environmental DNA to traditional field methods for regional-scale monitoring of an endangered aquatic species. *Molecular Ecology Resources*, 16, 895–908.
- Schmidt, B. R., Kery, M., Ursenbacher, S., Hyman, O. J., & Collins, J. P. (2013). Site occupancy models in the analysis of environmental DNA presence/absence surveys: A case study of an emerging amphibian pathogen. *Methods in Ecology and Evolution*, 4, 646–653.
- Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated—reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, 15, 1289–1303.
- Schnell, I. B., Sollmann, R., Calvignac-Spencer, S., Siddall, M. E., Douglas, W. Y., Wilting, A., & Gilbert, M. T. P. (2015). iDNA from terrestrial haematophagous leeches as a wildlife surveying and monitoring tool—prospects, pitfalls and avenues to be developed. *Frontiers in Zoology*, 12, 1.
- Schnell, I. B., Thomsen, P. F., Wilkinson, N., Rasmussen, M., Jensen, L. R., Willerslev, E., ... Gilbert, M. T. P. (2012). Screening mammal biodiversity using DNA from leeches. *Current Biology*, 22, R262–R263.
- Shaw, J. L., Clarke, L. J., Wedderburn, S. D., Barnes, T. C., Weyrich, L. S., & Cooper, A. (2016). Comparison of environmental DNA metabarcoding and conventional fish survey methods in a river system. *Biological Conservation*, 197, 131–138.
- Shelton, A. O., O'Donnell, J. L., Samhouri, J. F., Lowell, N., Williams, G. D., & Kelly, R. P. (2016). A framework for inferring biological communities from environmental DNA. *Ecological Applications*, 26, 1645–1659.
- Sigsgaard, E. E., Nielsen, I. B., Bach, S. S., Lorenzen, E. D., Robinson, D. P., Knudsen, S. W., ... Willerslev, E. (2016). Population characteristics of a large whale shark aggregation inferred from seawater environmental DNA. *Nature Ecology & Evolution*, 1, 0004.
- Simmons, M., Tucker, A., Chadderton, W. L., Jerde, C. L., & Mahon, A. R. (2015). Active and passive environmental DNA surveillance of aquatic invasive species. *Canadian Journal of Fisheries and Aquatic Sciences*, 73, 76–83.
- Sohlberg, E., Bomberg, M., Miettinen, H., Nyyssönen, M., Salavirta, H., Vikman, M., & Itävaara, M. (2015). Revealing the unexplored fungal communities in deep groundwater of crystalline bedrock fracture zones in Ouliluoto, Finland. *Frontiers in Microbiology*, 6, 573.
- Somervuo, P., Koskela, S., Pennanen, J., Henrik Nilsson, R., & Ovaskainen, O. (2016). Unbiased probabilistic taxonomic classification for DNA barcoding. *Bioinformatics*, 32, 2920–2927.
- Somervuo, P., Yu, D. W., Xu, C. C., Ji, Y., Hultman, J., Wirta, H., & Ovaskainen, O. (2017). Quantifying uncertainty of taxonomic placement in DNA barcoding and metabarcoding. *Methods in Ecology and Evolution*, 8, 398–407.
- Sommeria-Klein, G., Zinger, L., Taberlet, P., Coissac, E., & Chave, J. (2016). Inferring neutral biodiversity parameters using environmental DNA data sets. *Scientific Reports*, 6, 35644.
- Sønstebo, J., Gielly, L., Brysting, A., Elven, R., Edwards, M., Haile, J., ... Sannier, J. (2010). Using next-generation sequencing for molecular reconstruction of past Arctic vegetation and climate. *Molecular Ecology Resources*, 10, 1009–1018.
- Spens, J., Evans, A. R., Halfmaerten, D., Knudsen, S. W., Sengupta, M. E., Mak, S. S., ... Hellström, M. (2017). Comparison of capture and storage methods for aqueous microbial eDNA using an optimized extraction protocol: Advantage of enclosed filter. *Methods in Ecology and Evolution*, 8, 635–645.
- Stankovic, S., Kalaba, P., & Stankovic, A. R. (2014). Biota as toxic metal indicators. *Environmental Chemistry Letters*, 12, 63–84.
- Stoeckle, M. Y., Soboleva, L., & Charlop-Powers, Z. (2017). Aquatic environmental DNA detects seasonal fish abundance and habitat preference in an urban estuary. *PLoS ONE*, 12, e0175186.
- Stribling, J. B., Pavlik, K. L., Holdsworth, S. M., & Leppo, E. W. (2008). Data quality, performance, and uncertainty in taxonomic identification for biological assessments. *Journal of the North American Benthological Society*, 27, 906–919.
- Taberlet, P., Coissac, E., Hajibabaei, M., & Rieseberg, L. H. (2012). Environmental DNA. *Molecular Ecology*, 21, 1789–1793.
- Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding. *Molecular Ecology*, 21, 2045–2050.
- Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., ... Willerslev, E. (2007). Power and limitations of the chloroplast trn L (UAA) intron for plant DNA barcoding. *Nucleic Acids Research*, 35, e14–e14.
- Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., ... Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies. *Molecular Ecology*, 21, 1816–1820.
- Takahara, T., Minamoto, T., & Doi, H. (2013). Using environmental DNA to estimate the distribution of an invasive fish species in ponds. *PLoS ONE*, 8, e56584.
- Tang, C. Q., Humphreys, A. M., Fontaneto, D., & Barraclough, T. G. (2014). Effects of phylogenetic reconstruction method on the robustness of species delimitation using single-locus data. *Methods in Ecology and Evolution*, 5, 1086–1094.
- Taylor, H.R., & Gemmell, N.J. (2016). Emerging Technologies to Conserve Biodiversity: Further Opportunities via Genomics. Response to Pimm et al. *Trends in Ecology & Evolution*, 31, 171–172.
- Tedersoo, L., Bahram, M., Pölme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., ... Suija, A. (2014). Global diversity and geography of soil fungi. *Science*, 346, 1256688.
- Thomsen, P. F., Kielgast, J., Iversen, L. L., Møller, P. R., Rasmussen, M., & Willerslev, E. (2012). Detection of a diverse marine fish fauna using environmental DNA from seawater samples. *PLoS ONE*, 7, e41732.
- Thomsen, P., Kielgast, J., Iversen, L. L., Wiuf, C., Rasmussen, M., Gilbert, M. T. P., ... Willerslev, E. (2012). Monitoring endangered freshwater biodiversity using environmental DNA. *Molecular Ecology*, 21, 2565–2573.
- Thomsen, P. F., Møller, P. R., Sigsgaard, E. E., Knudsen, S. W., Jørgensen, O. A., & Willerslev, E. (2016). Environmental DNA from Seawater Samples Correlate with Trawl Catches of Subarctic, Deepwater Fishes. *PLoS ONE*, 11, e0165252.
- Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA—an emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, 183, 4–18.
- Torti, A., Lever, M. A., & Jørgensen, B. B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Marine Genomics*, 24, 185–196.
- Tréguier, A., Paillisson, J. M., Dejean, T., Valentini, A., Schlaepfer, M. A., & Roussel, J. M. (2014). Environmental DNA surveillance for invertebrate species: Advantages and technical limitations to detect invasive crayfish *Procambarus clarkii* in freshwater ponds. *Journal of Applied Ecology*, 51, 871–879.
- Turner, C. R., Barnes, M. A., Xu, C. C., Jones, S. E., Jerde, C. L., & Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous microbial eDNA. *Methods in Ecology and Evolution*, 5, 676–684.
- Turner, C. R., Uy, K. L., & Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, 183, 93–102.
- Valentini, A., Pompanon, F., & Taberlet, P. (2009). DNA barcoding for ecologists. *Trends in Ecology & Evolution*, 24, 110–117.
- Valentini, A., Taberlet, P., Miaud, C., Civade, R., Herder, J., Thomsen, P. F., ... Boyer, F. (2016). Next-generation monitoring of aquatic



- biodiversity using environmental DNA metabarcoding. *Molecular Ecology*, 25, 929–942.
- Valiere, N., & Taberlet, P. (2000). Urine collected in the field as a source of DNA for species and individual identification. *Molecular Ecology*, 9, 2150–2152.
- Vörös, J., Márton, O., Schmidt, B. R., Gál, J. T., & Jelić, D. (2017). Surveying Europe's Only Cave-Dwelling Chordate Species (*Proteus anguinus*) Using Environmental DNA. *PLoS ONE*, 12, e0170945.
- Vörösmarty, C. J., McIntyre, P. B., Gessner, M. O., Dudgeon, D., Prusevich, A., Green, P., ... Liermann, C. R. (2010). Global threats to human water security and river biodiversity. *Nature*, 467, 555–561.
- Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, 73, 5261–5267.
- West, J. S., Atkins, S. D., Emberlin, J., & Fitt, B. D. (2008). PCR to predict risk of airborne disease. *Trends in Microbiology*, 16, 380–387.
- Wheeler, Q. D., Raven, P. H., & Wilson, E. O. (2004). Taxonomy: Impediment or expedient? *Science*, 303, 285–285.
- Wilcox, T. M., McKelvey, K. S., Young, M. K., Lowe, W. H., & Schwartz, M. K. (2015). Environmental DNA particle size distribution from Brook Trout (*Salvelinus fontinalis*). *Conservation Genetics Resources*, 7, 639–641.
- Willerslev, E., Cappellini, E., Boomsma, W., Nielsen, R., Hebsgaard, M. B., Brand, T. B., ... Dahl-Jensen, D. (2007). Ancient biomolecules from deep ice cores reveal a forested southern Greenland. *Science*, 317, 111–114.
- Willerslev, E., Davison, J., Moora, M., Zobel, M., Coissac, E., Edwards, M. E., ... Haile, J. (2014). Fifty thousand years of Arctic vegetation and megafaunal diet. *Nature*, 506, 47–51.
- Willerslev, E., Hansen, A. J., Binladen, J., Brand, T. B., Gilbert, M. T. P., Shapiro, B., ... Cooper, A. (2003). Diverse plant and animal genetic records from Holocene and Pleistocene sediments. *Science*, 300, 791–795.
- Willerslev, E., Hansen, A. J., Christensen, B., Steffensen, J. P., & Arcander, P. (1999). Diversity of Holocene life forms in fossil glacier ice. *Proceedings of the National Academy of Sciences of the United States of America*, 96, 8017–8021.
- Willoughby, J. R., Wijayawardena, B. K., Sundaram, M., Swihart, R. K., & DeWoody, J. A. (2016). The importance of including imperfect detection models in eDNA experimental design. *Molecular Ecology Resources*, 16, 837–844.
- Wood, J. R., Wilmshurst, J. M., Wagstaff, S. J., Worthy, T. H., Rawlence, N. J., & Cooper, A. (2012). High-resolution coproecology: Using coprolites to reconstruct the habits and habitats of New Zealand's extinct upland moa (*Megalapteryx didinus*). *PLoS ONE*, 7, e40025.
- Xu, C. C., Yen, I. J., Bowman, D., & Turner, C. R. (2015). Spider web DNA: A new spin on noninvasive genetics of predator and prey. *PLoS ONE*, 10, e0142503.
- Yamamoto, S., Masuda, R., Sato, Y., Sado, T., Araki, H., Kondoh, M., ... Miya, M. (2017). Environmental DNA metabarcoding reveals local fish communities in a species-rich coastal sea. *Scientific Reports*, 7, 40368.
- Yamamoto, S., Minami, K., Fukaya, K., Takahashi, K., Sawada, H., Murakami, H., ... Kondoh, M. (2016). Environmental DNA as a 'Snapshot' of Fish Distribution: A Case Study of Japanese Jack Mackerel in Maizuru Bay, Sea of Japan. *PLoS ONE*, 11, e0149786.
- Yilmaz, P., Kottmann, R., Field, D., Knight, R., Cole, J. R., Amaral-Zettler, L., ... Cochrane, G. (2011). Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIXS) specifications. *Nature Biotechnology*, 29, 415–420.
- Yoccoz, N., Bräthen, K., Gjelty, L., Haile, J., Edwards, M., Goslar, T., ... Pompanon, F. (2012). DNA from soil mirrors plant taxonomic and growth form diversity. *Molecular Ecology*, 21, 3647–3655.
- Zaiko, A., Martinez, J. L., Schmidt-Petersen, J., Ribicic, D., Samuiloviene, A., & Garcia-Vazquez, E. (2015). Metabarcoding approach for the ballast water surveillance—An advantageous solution or an awkward challenge? *Marine Pollution Bulletin*, 92, 25–34.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

**How to cite this article:** Deiner K, Bik HM, Mächler E, et al. Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Mol Ecol*. 2017;26:5872–5895. <https://doi.org/10.1111/mec.14350>